

ESSAYS IN APPLIED BEHAVIORAL MICROECONOMICS

GIOVANNI PACI

Submitted in partial fulfilment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

© 2014

GIOVANNI PACI

All Rights Reserved

ABSTRACT

ESSAYS IN APPLIED BEHAVIORAL MICROECONOMICS

GIOVANNI PACI

Cognitive and emotional factors have played a larger role in economists' understanding of the world in the last decades. While earlier work has focused on experimental and theoretical results, a larger number of recent contributions have tested ideas from the field of Psychology using econometric methods for causal identification on field data. This line of research seeks to analyze market situations in which specific psychological factors can be identified to cause observed economic behavior. My dissertation, at the intersection of Behavioral and Applied Microeconomics, offers examples of behavior in which cognitive aspects are shown to play a central role and is unified across the three chapters by a common methodological approach.

The first chapter, based on joint work with Kareem Haggag, reports evidence from tipping behavior of New York city taxicab customers. For credit card payments, the payment screen in the car displays suggested tip amounts. In particular, for one of the main companies, the suggested amounts are \$2, \$3, \$4 for fares below \$15, and 20, 25 or 30 percent above \$15. Using this variation, the chapter shows that suggestions play an important role in tipping behavior of customers: comparing rides below and above \$15 using regression discontinuity methods, it is possible to show a large local causal effect of the suggestions on average tips. Moreover, a backlash effect is observed, as more customers decide not to tip on a credit card at all. These findings contribute to our understanding of default effects beyond the area of tipping, for instance in savings. An even broader lesson is that these findings isolate a case in which cognitive and emotional responses are likely to mediate the relationship between preferences and choice.

The second chapter, based on a joint work with Kareem Haggag, presents field evidence

on cheating behavior. During the two years 2008-2010, several taxi drivers cheated customers by charging a higher fare amount that is allowed only for rides outside the city even for rides in the city. The choice of whether to cheat a customer on a individual ride is shown to be affected by loss aversion. The estimates can be effectively reconciled by models of reference-dependent preferences that take drivers' expectation as reference points: drivers are more likely to cheat on those rides within a shift in which they are below expectations. The results highlight the role played by a classic decision-making bias in shaping unethical behavior in a market. These findings suggests that cognitive and emotional aspects of the valuation of benefits are relevant to our economic understanding of ethical problems.

The third chapter presents regression-discontinuity evidence on an investment-incentive program. The methodology, which compares firms who received the award with those that marginally lost it, allows for a cleaner identification of the effect of the policy. In this last essay, the conceptual tools from Applied Microeconomics used in the first chapter are put to work in the context of firms' behavior. The tool allows one to show in a straightforward manner the main outcomes of the policy.

Table of Contents

| | |
|--|-------------|
| List of Tables | iv |
| List of Figures | vi |
| Acknowledgements | viii |
| 1 Default Tips | 1 |
| 1.1 Introduction | 2 |
| 1.2 Institutional Context and Data | 6 |
| 1.2.1 Institutional Context | 6 |
| 1.2.2 Data Description | 8 |
| 1.3 Regression Discontinuity | 11 |
| 1.3.1 Visual Evidence | 11 |
| 1.3.2 Regression | 11 |
| 1.3.3 Other Outcomes of Interest | 15 |
| 1.3.4 Heterogeneity | 18 |
| 1.4 Comparing Across Vendors | 19 |
| 1.5 Conclusion | 25 |

| | | |
|----------|--|-----------|
| 2 | Reference Dependent Fraud | 28 |
| 2.1 | Introduction | 29 |
| 2.2 | Conceptual Framework | 36 |
| 2.3 | Data Description and Variable Construction | 40 |
| 2.4 | Empirical Strategy | 44 |
| 2.4.1 | Shift-Level Results | 44 |
| 2.4.2 | Construction of the Income Gap Measures | 51 |
| 2.4.3 | Identification and Instrumental Variable Analysis | 55 |
| 2.5 | Results | 57 |
| 2.5.1 | Summary Statistics | 57 |
| 2.5.2 | Linear Probability Model and Conditional Logit Model Estimates . . | 61 |
| 2.5.3 | Instrumental Variable Estimates | 67 |
| 2.5.4 | Liquidity Constraints | 71 |
| 2.6 | Conclusion | 73 |
| 3 | Investment Incentives for Lagging Areas: Evaluating Firms' Responses to Incentive Subsidies using a Regression Discontinuity Approach | 76 |
| 3.1 | Introduction | 77 |
| 3.2 | Institutional Context and Data | 82 |
| 3.2.1 | Institutional Context | 82 |
| 3.2.2 | Data Description | 84 |
| 3.3 | Regression Discontinuity | 87 |
| 3.3.1 | Methodology | 87 |
| 3.3.2 | Visual Evidence | 88 |
| 3.3.3 | Regression Analysis | 90 |

| | | |
|-------|--|------------|
| 3.4 | Heterogeneity | 93 |
| 3.5 | Conclusion | 96 |
| | Bibliography | 97 |
| | Appendices | 107 |
| | Appendix A Appendix for Chapter 1 | 108 |
| A.1 | Passenger Information Monitor Screen Examples | 108 |
| A.2 | Supplement to Section 2 (Regression Discontinuity) | 108 |
| A.3 | Supplement to Section 3 (Comparing Across Vendors) | 111 |
| A.4 | Data Cleaning | 115 |
| | Appendix B Appendix for Chapter 2 | 117 |
| B.1 | (A)Data Cleaning | 117 |
| B.2 | (B)Alternative Fraud Definitions | 118 |
| B.3 | (C)Robustness | 124 |
| B.3.1 | Alternative Income Gaps | 124 |
| B.3.2 | Autocorrelation | 129 |
| B.3.3 | Fixed effects contributions to Shift Income and Frauds | 129 |
| B.3.4 | JFK/LGA regression table | 132 |
| | Appendix C Appendix for Chapter 3 | 134 |
| C.1 | Summary Statistics | 134 |
| C.2 | Graphical evidence for other Outcome Variables | 136 |

List of Tables

| | | |
|-----|---|-----|
| 1.1 | Summary Statistics by Ride | 10 |
| 1.2 | Regression Discontinuity Estimates of the Effect on Tip Amount | 15 |
| 1.3 | Regression Discontinuity Estimates of the Effect on Alternative Outcomes | 16 |
| 1.4 | OLS - Comparison of Vendor (20%/25%/30%) and Competitor (15%/20%/25%) | 24 |
| 2.1 | OLS: Proportion of Rides with Fraud, by Shift | 47 |
| 2.2 | Summary Statistics - Ride level | 59 |
| 2.3 | Linear Probability Model: Fraud Probability, by Ride | 63 |
| 2.4 | Conditional Logit Model: Fraud Probability, by Ride | 67 |
| 2.5 | IV: 2SLS Estimates and First-Stage: Fraud Probability, by Ride | 69 |
| 2.6 | LPM Evidence on Liquidity Constraints: Fraud Probability, by Ride | 73 |
| 3.1 | Mean awards by program round | 86 |
| 3.2 | Regression discontinuity estimates for the main outcomes of interest, by pre- and post-award year | 92 |
| 3.3 | Regression discontinuity estimates, post- pre-award differences | 93 |
| 3.4 | Regression discontinuity estimates, post- pre-award differences by geography | 95 |
| A.1 | Regression discontinuity for trip distance, ride duration, hour of pick-up, day of the week, and passenger count | 110 |

| | | |
|------|--|-----|
| A.2 | Heterogeneity by Number of Passengers: Regression Discontinuity estimates of Default Effect on Tip Amount | 111 |
| A.3 | OLS - Comparison of Vendor (20%/25%/30%) and Competitor (15%/20%/25%) | 114 |
| B.1 | Fraud Summary Stats | 120 |
| B.2 | Other Fraud Definitions - Summary Stats | 123 |
| B.3 | Other Fraud Definitions - LPM | 124 |
| B.4 | Summary Statistics: Alternative Income Gap Measures | 126 |
| B.5 | LPM for Alternative Income Gap Measures | 127 |
| B.6 | LPM for Polynomial Income Gap Measures | 128 |
| B.7 | LPM with income inclusive frauds - Panel B | 129 |
| B.8 | LPM with income inclusive frauds - Panel B | 130 |
| B.9 | LPM with income inclusive frauds - Panel B | 131 |
| B.10 | IV: 2SLS Estimates and First-Stage (JFK) | 133 |
| C.1 | Summary statistics by program round and year | 135 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | Default Tip Suggestions by Vendor and Competitor | 7 |
| 1.2 | Histogram of Fares | 12 |
| 1.3 | Lowess Smoothed Mean Tip Percentages Within Each Discrete Fare Amount (\$0.40 Intervals) | 13 |
| 1.4 | Lowess Smoothed Mean Tip Percentages Within Each Discrete Fare Amount (\$0.40 Intervals), for Vendor and Competitor | 14 |
| 1.5 | Lowess Smoothed Mean of “Zero-Valued Tip” Within Each Discrete Fare Amount (\$0.40 Intervals) | 17 |
| 1.6 | Coefficients from Regression Discontinuity Estimates in Income Quantile Sub- Samples | 20 |
| 1.7 | Distribution of Fares (Panel A) and Tip Percentages (Panel B) across Vendor and Competitor Equipped Taxis | 22 |
| 2.1 | An example of Rate Code for activation | 43 |
| 2.2 | Crawford and Meng’s Distance from Expectation Measures: Proportion of Rides with Fraud, by Shift | 49 |
| 2.3 | Average Fraud by Income Gap measure and Trip Number within a Shift . . | 51 |
| 2.4 | Empirical Fraud Probability by Income Gap measure | 61 |
| 2.5 | Distribution of Marginal Effect on Positive Income Gap, By Driver | 65 |

| | | |
|-----|---|-----|
| 3.1 | Histogram of relative rank | 89 |
| 3.2 | Local polynomial regressions for investment/sales for three years before and after the award | 90 |
| A.1 | Passenger Information Monitors (PIM): Vendor (Top) & Competitor (Bottom) | 109 |
| A.2 | Vendor PIM Examples (above and below \$15) | 110 |
| A.3 | Proportion of Rides Originating with a Vendor versus a Competitor Equipped Cab, By census tract of Pick-Up Location | 112 |
| A.4 | Distribution of Fares (A & C) and Tip Percentages (B & D) across Vendor and Competitor Equipped Taxis in JFK Sample (A & B) or JFK & LGA Sample (C & D) | 113 |
| B.1 | Alternative Income Gap Measures | 125 |
| C.1 | Local polynomial regressions for profit/sales for three years before and after the award | 136 |
| C.2 | Local polynomial regressions for labor costs/sales for three years before and after the award | 137 |

Acknowledgements

I would like to sincerely thank Supreet Kaur, Stephan Meier, Kiki Pop-Eleches, Bernard Salanié, and Eric Verhoogen. I would also like to thank my colleague Kareem Haggag, with whom I worked on a daily basis on several articles in the last years.

I also want to thank many faculty members and colleagues for their great advice and guidance over the course of my graduate career: Colin Camerer, Alessandra Casella, Stefano Della Vigna, Prajit K. Dutta, Susan Elmes, Ray Fisman, Adam Galinsky, Paul Glimcher, Eric Johnson, Navin Kartik, Todd Kumler, David Laibson, Ben Marx, Brian McManus, Massimo Morelli, Sendhil Mullainathan, Donald Ngwe, Brendan O’Flaherty, Petra Persson, Herakles Polemarchakis, Devin Pope, Suresh Naidu, Matthew Rabin, Evan Riehl, Matthew Shum, Paolo Siconolfi, Dick Thaler, Sebastien Turban, Miguel Urquiola and Michael Woodford.

Chapter 1

Default Tips

Giovanni Paci ¹

¹This chapter is based on joint work with Kareem Haggag and it is forthcoming on the *American Economic Journal - Applied Economics*.

1.1 Introduction

The large effects of default options on consumer choices have been documented in various high-stakes, but low-frequency contexts, ranging from organ donation to 401(k) contributions. Because defaults preserve freedom of choice, but nonetheless appear to strongly influence behavior, they have been of great interest to both policy makers and academics (Thaler and Sunstein 2008). In contrast to the extant literature, we study the effects of defaults on a frequently encountered consumer choice: the decision of how much to tip a service provider.² By studying tipping, we demonstrate the ability of defaults to nudge behavior in a decision problem which agents have arguably encountered enough times to learn their optimal responses. In doing so, we also extend the literature by documenting a case in which default effects were exploited by a for-profit industry.

Our study introduces a unique data set that contains fare information for 170 million NYC taxi rides over the calendar year of 2009. Among these rides, we have tip information for the 38 million credit card transactions, from which we use a sample of more than 13 million rides to study tipping (all credit card transactions on rides with no tolls, taxes, or surcharges). At the end of each ride, customers who used credit cards were presented with a screen that provided them with the option to either type in a desired tip amount or to press one of three buttons with default tip suggestions. During the period of study, one of the credit card machine companies offered different tip suggestions depending on whether the fare was above or below \$15. For rides under \$15, tip suggestions were \$2, \$3, and \$4, while rides above \$15 were presented with 20%, 25%, and 30% tip suggestions and the corresponding dollar amounts (\$3, \$3.75, \$4.5 for a \$15 fare). At the discontinuity, this shift represents an increase in the suggestion categories (low, medium, and high) of approximately

²Though difficult to precisely measure, Azar (2011) estimates total annual tipping in the US food industry alone at \$47 billion, or approximately 0.3% of annual GDP.

\$1, \$0.75, and \$0.50. Importantly, the shift in suggestions did not change the choice set; customers were still free to key in any tip amount. Under the assumption that all ride characteristics that affect tips vary smoothly with the base amount, the difference at the discontinuity can be used to identify the causal effect of this particular increase in default suggestions on tipping. We find that this local treatment effect is an increase in tip amounts of approximately \$0.27 - \$0.30, which is more than a 10% increase in the average tip at that margin.

The design of this choice context is not typical of the default effects literature. In order to complete a transaction, customers had to enter a tip amount. As such, our study is closely related to the literature on *active choice*, a type of design in which customers are not provided with a no-action option, and are thus required to actively choose among a set of similarly presented options (e.g. Carroll et al. (2009) in the context of retirement savings). Keller et al. (2011) go a step further, presenting cases in which some options are advantaged by using favorable language, a design they describe as *enhanced active choice*. Our context goes even further in the direction of a default, with the featured amounts being strongly effort advantaged and highly salient. These features draw the influence of these options over choice closer to that of a default, and so we describe the buttons as *default tip suggestions*.

To examine the role of default suggestions across a larger range of fares, we present a second econometric strategy. We use the quasi-random assignment of passengers to taxi cabs at LaGuardia airport to compare across credit card machine companies. For rides above \$15, both companies provided percentage defaults; however, one company provided 15%, 20%, and 25% percent, while the other provided defaults of 20%, 25%, and 30% percent. The distribution of tips clearly reflects this shift, and again, we find that higher defaults are associated with higher average tip amounts, controlling for time-invariant driver characteristics.

Having demonstrated the benefits of higher default suggestions on the intensive margin of tipping, we next highlight a potential cost of setting defaults too high. First, in both the regression discontinuity design and the comparison across vendors, we find that the higher default suggestions reduce the probability of leaving a tip that corresponds to one of the default suggestions (24 and 7.8 percentage point reductions respectively).³ More striking is the result that rides with the higher tip suggestions are over 50% more likely to receive a zero-valued tip than those with the lower suggestions (1.7 and 2.8 percentage point increases). Such customers may have been penalizing drivers for using tip defaults that are perceived as *unfairly* high. Similarly, the absence of the lower amount button may have been construed by customers as an attempt to manipulate their behavior which induced them to respond adversely, as described in social psychology by *reactance* theory (Brehm 1966).

Several factors may explain our observed default effects. Customers may be rationally inattentive, failing to compute their preferred tip due to the opportunity cost of time and/or the cognitive costs associated with that computation. For instance, customers may have difficulty converting between dollars and percentages, and the alternative measures may evoke different intuitions on the appropriate tip (Kahneman 2011, p.372).⁴ Moreover, customers that are unfamiliar with the tipping norm may interpret the defaults as the socially endorsed norm. Both uninformed and informed customers may experience disutility from deviating from these options. Ultimately, we cannot disentangle these different possible mechanisms.

³However, we also find that the average manually entered tip amount increases. Thus, it is not clear that those induced to leave a manual tip are leaving lower tips than they would with the lower suggestions.

⁴On its own, the difficulty of comparing across the measurements does not imply that we should find higher tip amounts for the percentage suggestions. One potential explanation of this pattern would rely on this computational difficulty interacting with a particular type of self-deception. If customers adhere to tip *percentage* norms, then dollar suggestions could result in less generous tips by lowering the cost of self-deception. For example, consider a customer that has a fare of \$13 and adheres to a 25% tipping norm (i.e. a tip of \$3.25). This customer may be able to convince herself that she is adhering to the norm by selecting the \$3 option (rounding in the direction of her self-interest), whereas she could not ignore her deviation from the norm if explicitly presented with the 25% option.

We build upon the broad literature on defaults. Default effects have been demonstrated across a wide variety of consumer choices. Most notably, Madrian and Shea (2001) and Choi, Laibson, and Madrian (2004) found large effects in retirement savings contributions, with Madrian and Shea (2001) finding a 50% increase in enrollment from switching from an opt-in to automatic enrollment default. In a similarly sparsely encountered consumer choice, Johnson and Goldstein (2003) and Abadie and Gay (2006) used cross-country analyses to suggest that presumed consent policies induce higher organ donation rates than opt-in policies. Johnson et al. (1993) studied a somewhat more frequently encountered type of decision problem, namely (car) insurance plan choice, while Johnson, Bellman, and Lohse (2002) studied default effects in the decision to accept email marketing. Our paper contributes by showing that default effects can persist in a similarly habitually encountered consumer choice, using a much larger naturalistic field data set. We also add to the literature by tracing out the response to higher defaults, including its limitations. Beshears et al. (2010) similarly study the limits to setting high defaults. They provide a case study of a firm that set the default contribution rate at 12%, a rate much higher than previously studied defaults in this area (2% - 6%) and one that the authors note was sub-optimal for all employees in the sample. They find that roughly 25% of employees remain at this default rate after 12 months of tenure, in comparison to the 60% adherence rate seen at firms in previous studies. In our study, we find that a substantial proportion are induced to opt out of the default when presented with the higher suggestions. We still find a higher average contribution despite this result; however, our analysis also highlights the emergence of a cost (zero-valued tips) that suggests a potential reduction in tips if defaults are set sufficiently high.

The paper proceeds as follows. Section 1.2 provides background on taxis and tipping and describes the data used. Section 1.3 presents our regression discontinuity results. Section 1.4 presents an analysis that compares across credit card machine companies. Section 1.5

concludes.

1.2 Institutional Context and Data

The data for our study were provided by the Taxi and Limousine Commission (TLC) of New York City. In May 2004, the TLC mandated that all taxi cabs be outfitted with a set of technological improvements, including the electronic collection and transmission of trip data and the introduction of equipment to accept credit cards.⁵ These technological improvements also marked the introduction of a system that measured and saved the GPS coordinates of all pick-up and drop-off locations. Though mandated in 2004, the entire taxi fleet was not outfitted with the equipment until 2008.⁶ Our data spans the entirety of 2009, covering all rides by licensed Yellow Cab drivers in NYC. Before describing the data, we first present details about the institutional context.

1.2.1 Institutional Context

During the period of study, three companies were contracted to provide taxi cabs with credit card machines. The largest two, which we denote as “Vendor” and “Competitor”, account for 49% and 45% of observations in the raw data respectively. Each taxi cab was equipped with its own Passenger Information Monitor (PIM) which would display advertisements and other viewing material during the ride. At the end of the ride, the PIM displayed a payment screen (see the Appendix for an example). Customers were presented with the base amount and had the option of keying in their own tip amount or using one of the suggested tip

⁵Source: http://www.nyc.gov/html/tlc/html/industry/taxicab_serv_enh.shtml.

⁶Source: “Despite some grumbling, however, the TLC is moving to install the devices in all cars by August 31.” <http://www.nysun.com/business/hot-tip-for-cabbies-credit-cards-boost-tips/72783>.

The standard city rate (Rate Code 1) charged customers \$2.50 upon entry, and \$0.40 for each additional unit.⁷ One unit is defined as either (1) a 60 second interval in which the car is idle or driving less than 6 miles per hour or (2) 0.20 miles when the car is driving 6 miles per hour or faster. Fractional amounts are rounded up to the next unit. Riders were also subject to different sets of surcharges depending on the period of the year or the time of the day. To maintain comparability on either side of the Vendor default discontinuity, we limit our analysis to a period of time during which there were no taxes or surcharges and we exclude rides with tolls. This period includes January 1, 2009 to October 31, 2009 (i.e. prior to the introduction of a \$0.50 MTA tax), and spans 6am to 4pm on Monday - Friday and 6am to 8pm on Saturday and Sunday (i.e. periods of time not subject to the \$1.00 peak weekday or \$0.50 night time surcharges). During this time period, the largest base amount to the left of the \$15 discontinuity was \$14.90, and the smallest base amount to the right of the discontinuity was \$15.30.

1.2.2 Data Description

Our preliminary data set included 170,896,479 observations. Though the TLC has its own private routine for removing electronic glitches, the provided data set still contained a number of possible electronic errors, including zero-valued distances/durations and surcharges that

⁷Other Rate Codes include:

- Rate Code 2 - Rides to and from JFK - Charged a flat rate of \$45.
- Rate Code 3 - Rides to and from Newark Airport - Charged the standard rate in addition to a \$15 surcharge.
- Rate Code 4 - Rides to Nassau or Westchester county - Charged the standard city rate while in city limits, and double the standard rate while in Nassau or Westchester county.
- Rate Code 5 - Rides outside NYC, excluding Nassau, Westchester, or Newark Airport - Charged flat rate (determined through negotiation between rider and driver).

Source: http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml.

did not correspond to the appropriate schedule. We took a number of steps to clean the data, which we outline in greater detail in the Appendix. Our largest sample reductions were the removal of Cash payments (approximately three-quarters of the sample), and the removal of rides with taxes or surcharges. Our final dataset consists of 13,820,784 rides. For the majority of our estimates, we limit our sample to rides complete on cars equipped by the Vendor (7.28 million) and to fares between \$5 and \$25 (6.22 million observations).

The variables provided in the data are as follows: anonymized driver identifier, car identifier, credit card machine company, payment type, ride duration, ride distance, number of passengers, fare, surcharge, MTA tax, toll amount, and pick-up and drop-off latitude, longitude, and time. Because we do not have an indicator for whether the customer physically selected one of the default suggestion buttons, we needed to create this key variable. To do so, we make the assumption that all tip amounts that correspond to one of the relevant tip suggestions (e.g. \$2/\$3/\$4 for Vendor if the base amount is less than \$15) were selected from one of these buttons. We thus make the assumption that customers recognize this congruence and save the time of keying in this amount by pressing a single button.

For the purpose of computing heterogeneous treatment effects, we use data from the American Community Survey’s 5 year estimates (2006 - 2010). This dataset provides census tract level summary statistics. We match these statistics to each pickup and drop-off location. To do so, we first assign each GPS coordinate to a census tract using a point-in-polygon operation in PostgreSQL (PostGIS). We then merge each pickup location and each drop-off location with the ACS census tract variables. We focus on one variable in particular: median household income. Table 1.1 provides summary statistics for the sample, split by Vendor and Competitor.

Table 1.1: Summary Statistics by Ride

| | (1) Competitor | (2) Vendor | (3) Difference [(1)-(2)] |
|---|-------------------|--------------------|--------------------------------|
| Fare | 9.690 (5.479) | 9.813 (5.668) | -0.123*** |
| Tip Amount | 1.694 (1.253) | 1.920 (1.431) | -0.227*** |
| Tip as Percentage of Fare | 18.266 (9.254) | 21.463 (16.602) | -3.196*** |
| Tip Corresponds to a Default Tip Option | 0.556 (0.497) | 0.495 (0.500) | 0.061*** |
| Ride Duration (Minutes; Dropoff Time - Pickup Time) | 12.674 (8.136) | 12.854 (7.931) | -0.180*** |
| Ride Distance (Miles) | 2.531 (2.310) | 2.596 (2.401) | -0.065*** |
| Zero Tip | 0.020 (0.139) | 0.029 (0.167) | -0.009*** |
| High Choice [Pr(Select 'High' Default Select a Default Option)] | 0.129 (0.335) | 0.037 (0.189) | 0.092*** |
| Med Choice [Pr(Select 'Med' Default Select a Default Option)] | 0.419 (0.493) | 0.184 (0.388) | 0.234*** |
| Low Choice [Pr(Select 'Low' Default Select a Default Option)] | 0.452 (0.498) | 0.778 (0.415) | -0.326*** |
| Observations | 6,542,807 | 7,277,977 | 13,820,784 |

Notes: Standard deviations are in parentheses. T-tests for the equality of columns 1 and 2 are rejected at the 1% level for all differences. The sample is limited to rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday). High Choice (Vendor: \$4 or 30%, Competitor: 25%), Medium Choice (V: \$3 or 25%, C: 20%), and Low Choice (V: \$2 or 20%, C: 15%) estimates are conditional on using a default tip option.

1.3 Regression Discontinuity

1.3.1 Visual Evidence

We start by presenting a visual analysis of the discontinuity. As a first test of the validity of the regression discontinuity approach, Figure 1.2 demonstrates that the density of the forcing variable is smooth. Next, we limit the forcing variable (fare) to be between \$5 and \$25 and calculate the mean tip percentage within each of the discrete fare amounts (\$0.40 increments). On each side of the discontinuity, we scatter plot these estimates and perform a lowess smoother separately on either side of the discontinuity. Figure 1.3 displays this plot for tip percentages on Vendor-equipped cabs, clearly demonstrating a jump at \$15. Finally, as a robustness test, Figure 1.4 repeats this graph for the Competitor, showing no jump at \$15.

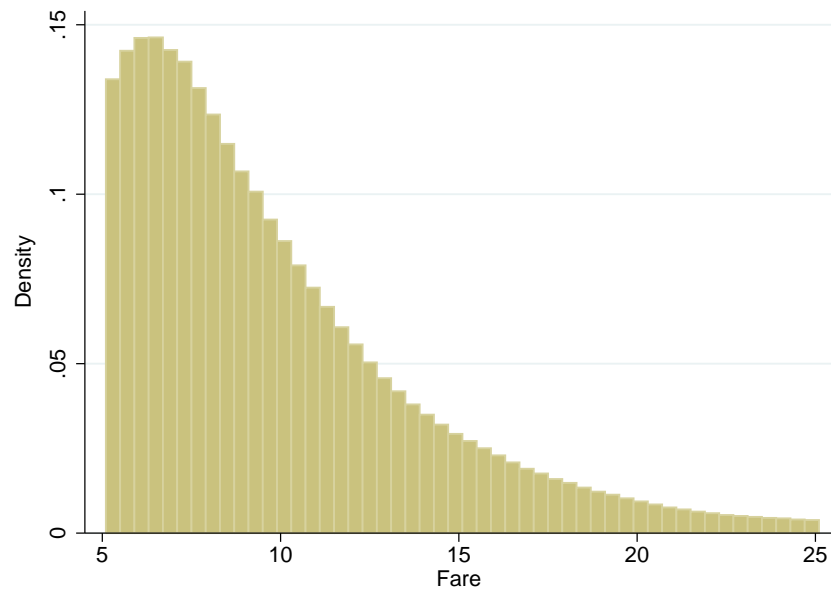
1.3.2 Regression

To supplement the visual evidence, we estimate a regression discontinuity model for the case of a forcing variable with discrete support. To the extent that the tip paid is smoothly related to the fare, observations at either side of the cutoff can be used to identify the causal effect of a change in the suggested amount. Following Lee and Card (2008), we estimate equation (1):

$$Y_r = \beta_0 \mathbf{1}(F_r \geq 15) + \beta_1 h(F_r - 15) + \beta_2 \mathbf{1}(F_r \geq 15) * g(F_r - 15) + \mathbf{X}_r \theta + u_r \quad (1.1)$$

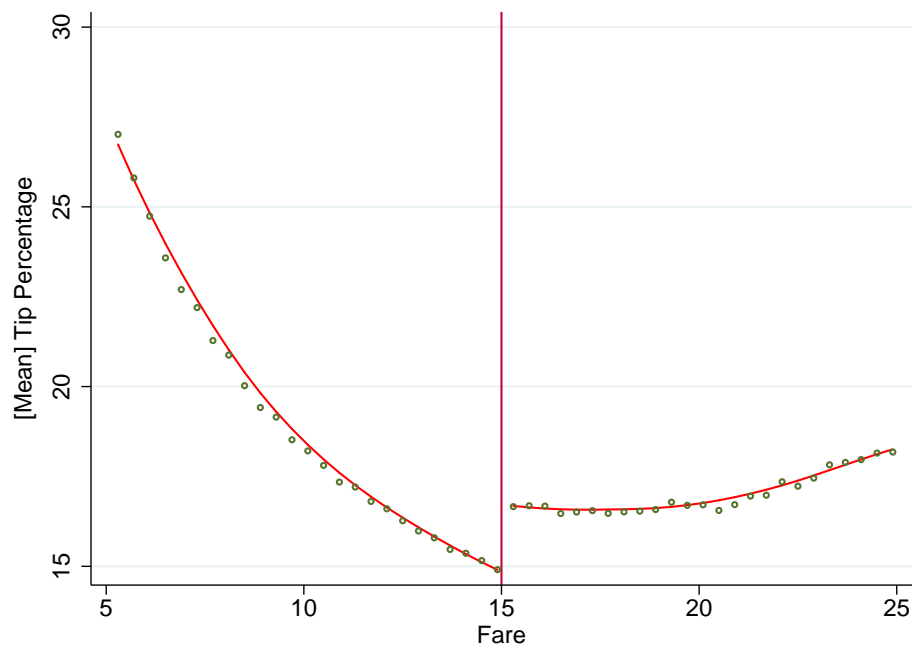
Where Y_r is the tip amount in dollars, $\mathbf{1}(F_r \geq 15)$ is an indicator function that the fare is greater than or equal to \$15, $h(F_r - 15)$ and $g(F_r - 15)$ are polynomials in the fare, centered at \$15, on either side of the discontinuity, and \mathbf{X}_r is a vector of fixed effects consisting of

Figure 1.2: Histogram of Fares



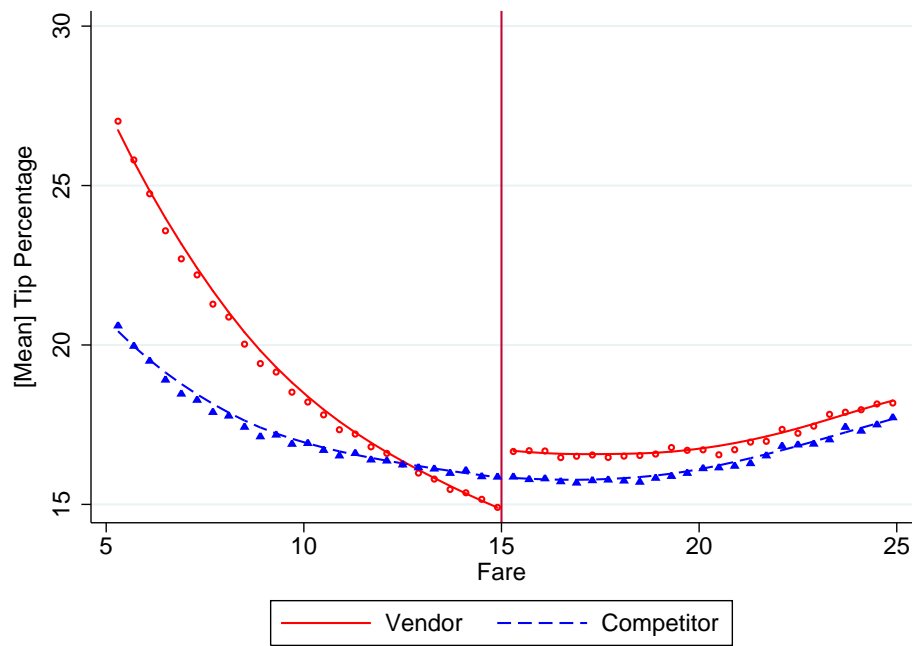
Notes: The sample is limited to fares between \$5 and \$25 on **Vendor-equipped** cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday). Fares are in \$0.40 intervals.

Figure 1.3: Lowess Smoothed Mean Tip Percentages Within Each Discrete Fare Amount (\$0.40 Intervals)



Notes: Each dot is the average within a discrete fare amount (\$0.40 intervals). Solid lines display the smoothed values from locally weighted regressions performed separately on either side of the discontinuity (\$15). The sample is limited to fares between \$5 and \$25 on **Vendor-equipped** cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday). Fares are in \$0.40 intervals.

Figure 1.4: Lowess Smoothed Mean Tip Percentages Within Each Discrete Fare Amount (\$0.40 Intervals), for Vendor and Competitor



Notes: Each dot is the average within a discrete fare amount (\$0.40 intervals). Solid and dashed lines display the smoothed values from locally weighted regressions performed separately on either side of the discontinuity (\$15). The sample is limited to fares between \$5 and \$25 on rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday). Fares are in \$0.40 intervals.

pick-up hour, day of week, pick-up borough, and drop-off borough indicators. Since our source of variation is at the fare value relative to the discontinuity, rather than the ride level, we follow Lee and Card (2008) and cluster our standard errors at the level of the forcing variable, thereby correcting our degrees of freedom and allowing for random specification errors due to the discrete bins. We estimate four specifications in Table 1.2, starting with a 2nd order polynomial in the first column up to a 5th order polynomial in the last column. Our local treatment effect is a \$0.27 to \$0.30 increase in tip amounts over a baseline level at the cut-off of \$2.22.

Table 1.2: Regression Discontinuity Estimates of the Effect on Tip Amount

| | Tip Amount | | | |
|------------------------------|---------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) |
| $\mathbb{1}(Fare_r \geq 15)$ | 0.292*** (0.004) | 0.276*** (0.006) | 0.275*** (0.008) | 0.296*** (0.010) |
| N | 6,218,196 | 6,218,196 | 6,218,196 | 6,218,196 |
| r ² | 0.207 | 0.207 | 0.207 | 0.207 |
| DepVarMean | 2.221 | 2.221 | 2.221 | 2.221 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at each fare value (\$0.40 intervals), in parentheses. Columns (1) - (4) present 2nd-5th order polynomials. $\mathbb{1}(Fare_r \geq 15)$ is an indicator function that the fare is greater than or equal to \$15. DepVarMean is the mean of the dependent variable on rides with fares of \$14.90. All specifications include fixed effects for driver, pick-up day of the week, pick-up hour, pick-up location borough, and drop-off location borough. The sample is limited to fares between \$5 and \$25 on Vendor-equipped cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday).

1.3.3 Other Outcomes of Interest

Our primary outcome of interest, tip amount, is produced through movements along an extensive margin (using a default suggestion) and two intensive margins (amounts tipped

either manually or through one of the suggestions). Table 1.3 presents results for a number of other variables.

Table 1.3: Regression Discontinuity Estimates of the Effect on Alternative Outcomes

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---------------------|---------------------|----------------------|---------------------|---------------------|----------------------|----------------------|---------------------|---------------------|
| | Tip | Default | Default | Manual | High | Med | Low | Zero |
| | Percent | Tip | Tip Amt | Tip Amt | Choice | Choice | Choice | Tip |
| $1(Fare_r \geq 15)$ | 2.025*** (0.038) | -0.243*** (0.002) | 0.714*** (0.003) | 0.368*** (0.011) | -0.021*** (0.001) | -0.169*** (0.002) | 0.190*** (0.003) | 0.017*** (0.001) |
| N | 6,218,196 | 6,218,196 | 3,227,726 | 2,990,470 | 3,227,726 | 3,227,726 | 3,227,726 | 6,218,196 |
| r ² | 0.097 | 0.058 | 0.568 | 0.122 | 0.015 | 0.029 | 0.038 | 0.008 |
| DepVarMean | 14.907 | 0.749 | 2.489 | 1.422 | 0.073 | 0.342 | 0.585 | 0.028 |

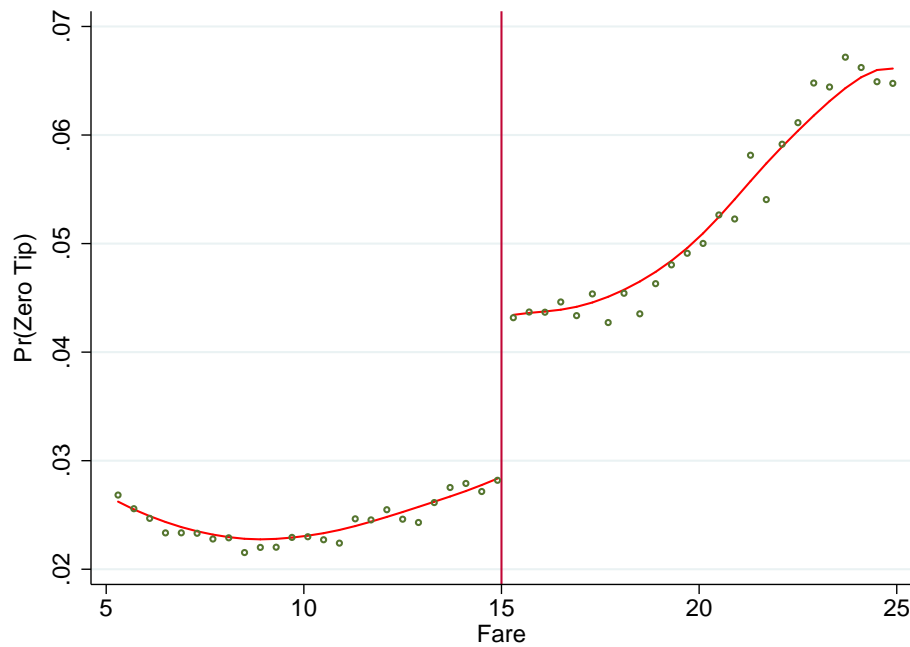
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at each fare value (\$0.40 intervals), in parentheses. $1(Fare_r \geq 15)$ is an indicator function that the fare is greater than or equal to \$15. Columns 1 (Tip Percent), 3 (Default Tip Amount), and 4 (Manual Tip Amount) use continuous outcome variables, while columns 2 (Default Tip), 5 (High Choice), 6 (Medium Choice), 7 (Low Choice), and 8 (Zero Tip) use binary outcome variables. The dependent variable in column 2 (Default Tip) takes on value 1 if the customer selected one of the default tip suggestions (buttons). The dependent variable in column 3 is the tip amount conditional on using one of the default tip suggestions. The dependent variable in column 4 is the tip amount conditional on manually entering a tip using the keypad. The High Choice (Vendor: \$4 or 30%, Competitor: 25%), Medium Choice (V: \$3 or 25%, C: 20%), and Low Choice (V: \$2 or 20%, C: 15%) variables are conditional on using a default tip option, and take on value 1 if the customer select the corresponding option. The dependent variable in column 8 (Zero Tip) takes on value 1 if the customer left zero credit card tip. DepVarMean is the mean of the dependent variable on rides with fares of \$14.90. All specifications use 3rd-order polynomials and include fixed effects for driver, pick-up day of the week, pick-up hour, pick-up location borough, and drop-off location borough. The sample is limited to fares between \$5 and \$25 on Vendor-equipped cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday).

Notably, the higher tip suggestions induce a 24 percentage point reduction in the probability of using a default suggestion, and a shift in the composition of those that use default suggestions toward the low option. However, since the low option is approximately equal to the medium option to the left of the discontinuity, we see a net increase in the amount tipped by those that select a default option. There is also an increase in amount tipped manually, which reflects a change in the composition of the tippers, but may also reflect the influence of the higher suggestions on those that would tip manually when faced with either set of suggestions.

Column 8 shows perhaps the most interesting of these behavioral responses. The probability of leaving no credit card tip increases by 1.7 percentage points when customers are faced with the higher tip suggestions. Figure 1.5 repeats the visual analysis for this outcome variable. This negative response is consistent with the social psychology literature on “reac-

Figure 1.5: Lowess Smoothed Mean of “Zero-Valued Tip” Within Each Discrete Fare Amount (\$0.40 Intervals)



*Notes: Each dot is the average within a discrete fare amount (\$0.40 intervals). Solid lines display the smoothed values from locally weighted regressions performed separately on either side of the discontinuity (\$15). The sample is limited to fares between \$5 and \$25 on **Vendor-equipped** cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday).*

tance”. Within this framework, individuals react to a perceived reduction in their freedom of choice by doing the opposite of the intended manipulation (Brehm 1966, 1989). Though customers’ choice sets of tip amounts were not changed by the shift to higher suggestions, they may have still perceived the replacement of the lower button amounts as a threat to

their behavioral freedom. The negative reaction also has some parallel in the vast literature on ultimatum games. Insofar as customers have some fixed notion of a “fair” tip, presenting the higher suggestions might have led them to punish the drivers by leaving a lower tip than would be provided in the absence of the “unfair” split. This result highlights a potential cost of setting defaults too high, although we cannot confirm whether this cost would persist in a context featuring homogeneous suggestions across vendors (e.g. all vendors currently offer the 20%/25%/30% suggestions). The backlash to high suggestions may hinge on the existence of a reference “fair tip” in a comparable market. Furthermore, without making strong assumptions, we cannot use this cost to trace out the set of optimal default suggestions. Finally, we cannot rule out an important possible confound. Since we do not observe cash tips, it may be possible that customers were induced to switch from paying both their tips and fares by credit to instead paying their fares by credit and their tips by cash.

1.3.4 Heterogeneity

We next examine heterogeneous treatment effects along proxied income. Because wealthier riders have a lower marginal utility of income, they have less incentive to be attentive to the shift in tip suggestions. Similarly, these customers will potentially have higher time costs. This reasoning suggests that higher income customers should (rationally) exhibit a larger default effect. Alternatively, wealthier individuals may have greater access to distraction-reducing devices, allowing them to deplete less attention during the day (Banerjee and Mulainathan 2008), and thus be less susceptible to default effects. A wealthier customer may also be more likely to take more taxi rides, and default effects possibly attenuate with experience (Lofgren et al. 2012). Ultimately, this exercise is theoretically ambiguous; however, it is an interesting source of potential heterogeneity that has been studied in other contexts. For example, Goldin and Homonoff (2012) found that low-income customers were more attentive

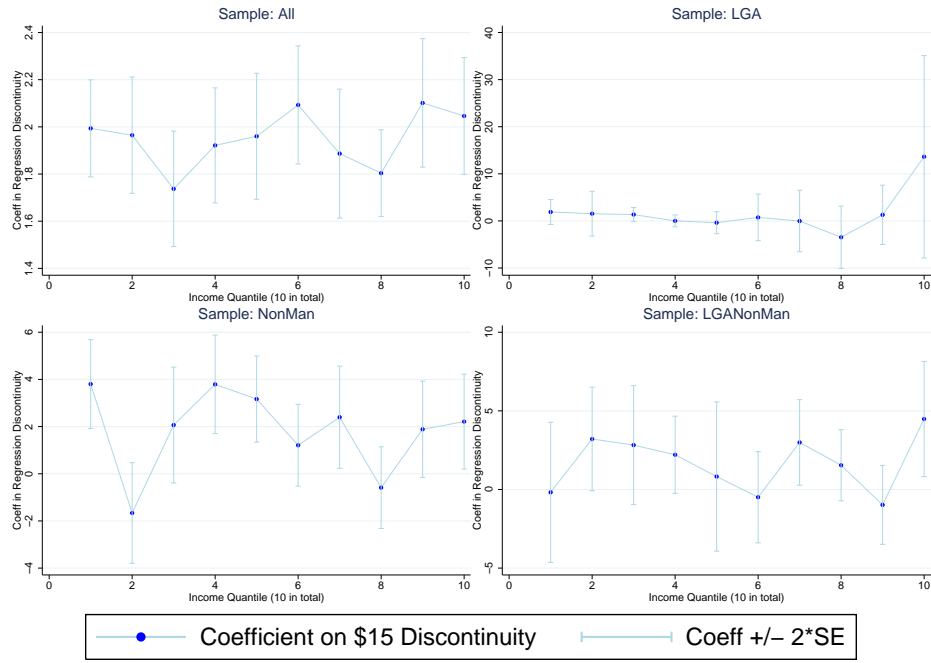
to a low salience cigarette tax than were high income customers. In contrast, Beshears et al. (2010) found that 401(k) savings defaults had a greater influence on low-income employees than on high-income employees.

To proxy income, we use a variety of different sample groups. For the full sample, we proxy customer income by the average of the median income associated with the pick up location and drop-off location census tracts. These pick-up and drop-off locations would not be an accurate assessment of tourists or any other customers that do not start or end at their home addresses. To partially reduce this concern, we use a variety of other specifications that attempt to remove these non-representative customers. One set of specifications limit the sample to rides that both start and end outside Manhattan. Another set of specifications limit to rides that either start or end at LaGuardia (LGA) airport, proxying income by the median income in the pick-up (drop-off) location census tract if the ride ends (starts) at LGA. Finally, the most conservative set of specifications limits to rides that start at LaGuardia airport and end outside Manhattan. We then split these rides into ten income quantiles and run the regression discontinuity for each of these sub-samples. Figure 1.6 plots the coefficients from these regressions, finding no systematic pattern in income. Although it might be that the absence of a clear pattern is due to measurement error in the income variable, the constancy in the discontinuity suggests that default effects are similar for customers across the different proxied income brackets in our sample.

1.4 Comparing Across Vendors

While our regression discontinuity design provides compelling identification, it is limited to a localized treatment effect. One way to expand upon our results would be to compare rides over which both credit card machine companies provided percentage default suggestions.

Figure 1.6: Coefficients from Regression Discontinuity Estimates in Income Quantile Sub-Samples



Notes: Robust standard errors clustered at each fare value (\$0.40 intervals). All specifications use 3rd-order polynomials and include fixed effects for driver, pick-up day of the week, pick-up hour, pick-up location borough, and drop-off location borough. The sample is limited to fares between \$5 and \$25 on Vendor-equipped cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday). All income measures correspond to American Community Survey 5-year (2006-2010) estimates of the census tract median income. The LGA sample is restricted to rides that either originate or end within the census tract associated with LaGuardia Airport. Income is proxied by the median income in the pick-up location census tract if the ride ends at LGA or by the income at the drop-off location if the ride starts at LGA. The NonMan sample is restricted to rides that both start and end outside Manhattan, and uses the average of the median income at the pick-up and drop-off census tracts. The LGANonMan sample corresponds to rides that start at LaGuardia Airport and end outside of Manhattan. The Average (Median ACS Census Tract) income within each quantile are as follows: **All**: 49,244 - 68,340 - 78,945 - 86,876 - 93,574 - 99,455 - 105,490 - 112,490 - 121,617 - 145,084. **NonMan**: 26,636 - 38,634 - 45,384 - 49,978 - 53,906 - 57,856 - 63,013 - 71,318 - 82,629 - 106,607. **LGA**: 27,775 - 43,948 - 52,174 - 58,943 - 69,240 - 81,628 - 92,334 - 102,506 - 114,219 - 138,099. **LGANonMan**: 28,271 - 42,718 - 48,115 - 51,752 - 54,076 - 57,739 - 61,705 - 69,823 - 77,697 - 118,845.

For fares above \$15, the Vendor provided suggestions of 20%, 25%, and 30%, while the competitor provided suggestions of 15%, 20%, and 25%. However, there are several potential differences in the matches of customers and drivers between the two companies (e.g. see the Appendix for a figure depicting the geospatial distribution of pick-up locations between the two companies). While we can control for the pick-up and drop-off location, there may be other unobservable differences in driver-customer match that affect the tip amounts. To address this challenge, we limit our analysis to rides that originate at LaGuardia airport.⁸ Customers queue at lines that contain a mix of taxis equipped with both credit card machine companies. Panel A of Figure 1.7 shows that the distribution of fares is comparable across the two credit card companies when we limit the sample to rides that are above \$15 and originate at LaGuardia.⁹

For the distribution of tip percentages, we limit the sample to fares with tip percentages less than or equal to 50% in order to provide greater visual clarity. Panel B of Figure 1.7 demonstrates the stark difference in the two distributions, with the higher set of defaults inducing a distribution that has significantly more density around its lowest option. The left tail of the figure is also larger for the Vendor; however, this effect is limited to zero-valued tips.

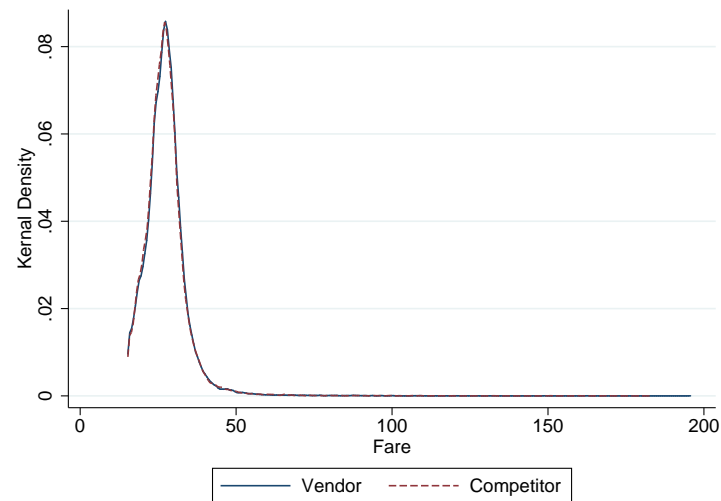
We provide a regression analysis of these effects in Table 1.4. To address concerns of any type of sorting between drivers and credit card machine companies, we also provide specifications with fixed effects for driver, pick-up hour, and borough of the drop-off location. These fixed effects specifications exploit the 7% of drivers in our sample that drove on cars

⁸We exclude JFK airport because the majority of rides use a \$45 fixed fare, complicating our placebo test of equality in fares between vendors. In the Appendix, we repeat the analysis in this section pooling both LGA and JFK observations. Our estimates in that pooled sample are qualitatively similar.

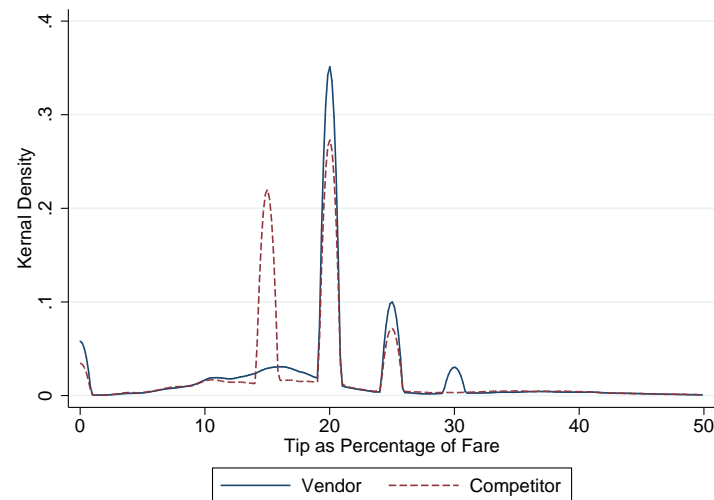
⁹However, it should be noted that a simple t-test of fare across the vendor (27.47) and competitor (27.36) rejects equality ($p = .0089$).

Figure 1.7: Distribution of Fares (Panel A) and Tip Percentages (Panel B) across Vendor and Competitor Equipped Taxis

[Panel A: Fare]



[Panel B: Tip Percentage]



Notes: The sample is limited to fares greater than \$15 on cab rides that originated at the census tract associated with LaGuardia Airport, without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday). Panel B is limited to rides with tip percentages less than 50%.

equipped by both companies on rides from LaGuardia Airport, allowing us to identify the coefficient on “Vendor” while controlling for time-invariant driver characteristics.¹⁰ However, it is important to note that we cannot control for the influence of possible additional differences between the two companies in the distribution of taxi cab characteristics, such as the display of the payment screens (e.g. the Vendor, unlike the Competitor, paired the percentage suggestions with their dollar conversions).

We find that the small difference in fare is significant at the 5% level in specifications that do not include fixed effects (Column 1), and insignificant in specifications that do include fixed effects (Column 2) – though these coefficients are not statistically different from each other. We find a significant increase in the tip percentage, consistent with our regression discontinuity estimates, though the effect size is smaller in magnitude. We also find a reduction in the probability of using one of the suggested amounts, also consistent with Section II. Columns 7 and 8 show a small decrease in the probability of leaving a tip greater than 0% but less than 10%; however, columns 9 and 10 show a significant increase in the probability of leaving a zero-valued tip. As with the regression discontinuity analysis, we cannot rule out the possibility that customers presented with the higher options were simply switching to paying their tips in cash. Unlike the regression discontinuity analysis presented in Section II, this comparison suffers from the potentially confounding influence of other differences between the two companies. For example, these zero-valued tip entries could reflect data errors that were more likely to be produced by Vendor machines. In cleaning the data, we removed all zero-valued distance and ride duration observations; however, we found that there were slightly more of these distance errors associated with the Competitor (0.88% vs. 0.65%) and more ride duration errors associated with the Vendor (.65% vs. .05%). It

¹⁰There may be time-varying characteristics that present a threat to identification, such as motivation (e.g. drivers may choose to use Vendor-equipped cabs on days in which they are more motivated to earn income).

Table 1.4: OLS - Comparison of Vendor (20%/25%/30%) and Competitor (15%/20%/25%)

| | Fare | | Tip Percent | | Default Tip | |
|----------------|---------------------|----------------------|---------------------|---------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Vendor | 0.128** (0.056) | 0.176 (0.146) | 0.675*** (0.099) | 0.699*** (0.259) | -0.083*** (0.004) | -0.078*** (0.013) |
| N | 100,577 | 100,577 | 100,577 | 100,577 | 100,577 | 100,577 |
| r2 | 0.000 | 0.295 | 0.001 | 0.265 | 0.007 | 0.222 |
| MeanDepVar | 27.047 | 27.047 | 18.652 | 18.652 | 0.634 | 0.634 |
| Fixed Effects? | | X | | X | | X |
| PVal_FEvSNoFE | | 0.721 | | 0.917 | | 0.661 |
| | TipPercent0to10 | | Zero Tip | | Tip25 | |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| Vendor | -0.003** (0.001) | -0.016*** (0.006) | 0.028*** (0.001) | 0.028*** (0.006) | 0.034*** (0.002) | 0.037*** (0.007) |
| N | 100,577 | 100,577 | 100,577 | 100,577 | 100,577 | 100,577 |
| r2 | 0.000 | 0.213 | 0.004 | 0.204 | 0.003 | 0.210 |
| MeanDepVar | 0.052 | 0.052 | 0.039 | 0.039 | 0.080 | 0.080 |
| Fixed Effects? | | X | | X | | X |
| PVal_FEvSNoFE | | 0.011 | | 0.998 | | 0.671 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the driver level, in parentheses. Even columns include fixed effects for driver, pick-up hour, and drop-off borough. The dependent variable in columns 5 and 6 (Default Tip) takes on value 1 if the customer selected one of the default tip suggestions (buttons). The dependent variable in columns 7 and 8 (Tip Percent > 0 & < 10) takes on value 1 if the tip is greater than 0% and less than 10% of the fare. The dependent variable in columns 9 and 10 (Zero Tip) takes on value 1 if the customer left zero credit card tip. The dependent variable in columns 11 and 12 (Tip Percent = 25) takes on value 1 if the customer selected the 25% tip button. DepVarMean is the mean of the dependent variable in the control group (rides on Competitor-equipped cabs). PVal_FEvsNoFE is the p-value from a Chow Test for the equality of coefficients across even and odd columns. The sample is limited to fares greater than \$15 on cab rides that originated at the census tract associated with LaGuardia Airport, without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday).

is possible that some of the zero-valued tip percentage entries are residual electronic errors or tests, and that these tip errors are more concentrated in Vendor credit card machines. Despite this issue, the results are congruent with the regression discontinuity results for the zero tipping outcome.

Finally, Columns 11 and 12 present the proportion of customers who selected the 25% button, i.e. the “high” option for Competitor cabs and the “middle” option for Vendor cabs. We find that the proportion of customers who select the 25% button increases by 3.7 percentage points when it is the “middle” option, relative to when it is the “high” option. This result is suggestive of customers being influenced by a context effect similar to the “compromise effect” (Simonson 1989). A typical test of the compromise effect would compare the tendency to choose an option (e.g. 25%) between a choice set in which it is the high option out of two (e.g. 20% and 25%) against a choice set that adds a higher option (e.g. 20%, 25%, and 30%).¹¹ Our test differs in that the addition of a higher button (30%) comes with the removal of a low button (15%). Nonetheless, this result is suggestive of a compromise effect operating, even when the choice set of possible tip amounts is preserved and only the set of default tip suggestions is manipulated.

1.5 Conclusion

Using an extensive dataset, we show that a small change in default tip suggestions has a significant effect on tipping amounts. Our data allow us to provide very clean identification

¹¹One could also examine the removal of an “inferior” option from the choice set (e.g. the removal of the 15% button) on the propensity to choose what was previously the middle option (20%). We do not find this to be a particularly compelling test of the compromise effect because 15% is a highly relevant (frequently chosen) alternative, rather than a decoy (in contrast to the rarely chosen 30% option). Thus, in that manipulation, the default effect is the more relevant phenomenon and appears to dominate the compromise effect (Figure 1.7b shows that the propensity to choose 20% actually increases when it becomes the low option).

in a large naturalistic field setting. We use a regression discontinuity design to show that an upward shift in the set of suggestions induces higher average tip amounts, despite significantly reducing the probability that customers use one of the defaults. To analyze consumers' responses across sets of suggestions that were closely framed in terms of percentages, as well as to provide less localized treatment effects, we performed a secondary analysis of trips originating at LaGuardia airport. Exploiting these quasi-random driver-to-customer matches, we again find that higher default tip suggestions (20%/25%/30% vs. 15%/20%/25%) result in higher average tip amounts. This analysis also reveals a potential cost of setting defaults too high – customers are also more likely to leave no tip in response to the higher defaults.

The observed default effect may be attributed to three potentially complementary mechanisms. First, customers may be rationally inattentive if the cognitive effort or time costs are sufficiently high to justify the additional tip.¹² A second possible explanation is that the default serves an information transmission purpose, acting as a signal of the social norm to unfamiliar customers.¹³ Finally, customers may experience disutility from deviating from the status quo, either due to social pressure or other forms of psychological resistance. Ultimately, our data do not allow us to cleanly parse the three proposed mechanisms.

¹²The effects of these changes in suggestions on overall spending are relatively small even for regular NYC taxi users, though sizable for taxi drivers. To perform a back-of-the-envelope calculation, we use a question about the frequency of taxi use from a voluntary passenger survey administered in 2010 by the Taxi and Limousine Commission (http://www.nyc.gov/html/tlc/downloads/pdf/tot_survey_results_02.10.11.pdf). We approximate the average of these bucketed responses to ~ 100 rides per year, and scale down by the proportion of rides paid by credit in 2009 ($\sim 25\%$). Even with this selected sample of passengers, if we extrapolate from our RDD ($\sim \$0.30$) or Across Vendor ($\sim \0.20) point estimates, the change in overall spending is just $\$7.25/\5 for a passenger who spends over $\$1,000$ a year on taxis. In contrast, for the median driver in the raw data with $\sim 5,000$ rides per year ($\sim 1,250$ by credit), similar calculations produce estimates of $\$375/\250 increases in their annual incomes. We should stress that extrapolating from our local average treatment effects requires making several strong and unrealistic assumptions – these calculations are reported here solely to give the reader a rough sense of magnitude.

¹³A prime example of this social pressure mechanism is expressed in a New York Sun magazine article on the introduction of the credit card system: *“It forces you to tip,” a Manhattan resident who recently tipped 15% on a \$14 fare, Greg Mack, said. “What if you didn’t enjoy the ride? It made me feel obligated.”* (Source: <http://www.nysun.com/business/hot-tip-for-cabbies-credit-cards-boost-tips/72783/>).

As firms begin to use the insights of behavioral economics to inform their product design and promotion, our study suggests that default effects can be exploited even in habitually encountered consumer choices. However, there may be a backlash to defaults that exceed certain thresholds, and so firms and policy makers alike should be cognizant of this potential cost when optimally designing their defaults.

Chapter 2

Reference Dependent Fraud

Giovanni Paci ¹

¹This chapter is based on joint work with Kareem Haggag.

2.1 Introduction

Prospect Theory has been used to explain otherwise puzzling behavior by agents as diverse as home sellers, professional golfers, stock market investors, capuchin monkeys, and students in the laboratory (Kahneman and Tversky 1979, Thaler et al. 1997, Odean 1998, Genesove and Mayer 2001, Chen et al. 2006, Pope and Schweitzer 2011). We add to this literature by presenting field evidence on the behavior of economic agents who commit crime in the workplace. We do so by studying the cheating decisions of New York City taxi drivers observed in field data over the course of one year. We find that deviations from expectations are among the considerations leading agents to commit fraud, and estimate that a driver who falls below his proxied income expectation becomes 7%-16% more likely to cheat a customer on a given ride (on a baseline cheating rate of 0.88 percent).

Using an instrumental variable approach that takes advantage of aggregate market shocks, we estimate an even larger response. This identification strategy allows us to credibly isolate deviations from expectations that are not generated by endogenous effort decisions by drivers. We do so by relying on the performance of other drivers operating in the same area and time span, similar to Camerer et al. (1997). Our results are in line with the growing evidence on the applicability of reference-dependent preferences for predicting behavior in many settings, including unethical behavior (Mas 2006, Card and Dahl 2012).

Our analysis is based on a real-world episode of wide-spread fraud. During the two year period 2008-2010, thousands of taxi drivers overcharged customers by applying a rate code that was intended to be reserved for rides outside the New York City limits.² The Taxi and Limousine Commission (TLC) uncovered this large-scale fraud scheme in 2010 by pairing the retrospective electronic trip sheet and GPS data, resulting in the incarceration of 59

²<http://www.nytimes.com/2010/03/13/nyregion/13taxi.html?pagewanted=all>.

drivers.³ We use the TLC data for the year 2009, which span over 170 million trips by more than 30,000 licensed yellow cab taxi drivers. Thus, we are able to identify when drivers overcharged their customers by applying a rate code that was only legally applicable for rides outside of Manhattan, using a similar methodology to the one adopted by the Taxi and Limousine Commission for the rides inside Manhattan. Moreover, we can track individual drivers' earnings from each ride during each labor spell (a shift). These combined features of the data provide us with a unique setting to investigate income-based reference-dependence as one of the potential determinants of cheating at the individual ride level.

Using all the rides for each driver within 2009, we estimate a within-driver measure of cumulative income expectations over each shift. For the purpose of identification, we exploit the fact that drivers' earnings are subject to randomness, as they receive a variable number of streets hails (depending on demand conditions) and these rides vary in their distance and duration. The unanticipated variation in search time and fares allows us to test whether drivers cheat more when their realized income is below our estimated expected income measure. We present evidence from linear probability models and from an instrumental variable approach. Ultimately, similar to the majority of the literature that estimates models containing expectation terms, our empirical approach tests the joint hypothesis that the expectation formation process is correctly specified and that loss aversion affects the probability of cheating. In particular, our specification requires that drivers continually evaluate their income stream relative to this running expectation, and that unexpected earnings fluctuations around the expectation are treated meaningfully. On the one hand, the expectation

³Manhattan District Attorney Cyrus R. Vance, Jr., announced the arrests of 59 New York City taxicab drivers for defrauding and stealing on September 21, 2010. Among the drivers, 45 were charged with two counts of Scheme to Defraud in the First Degree, a felony of class E, punishable with up to 4 years of prison. An additional 14 taxicab drivers were arrested and individually charged with a cumulative total of more than 5,000 counts of "petit larceny". The arrests followed an investigation run by the District Attorney and the Taxi and Limousine Commission. *The New York County District Attorney Office*, (2010).

formation process we postulate is admittedly a noisy measure of expectations at any given time, on the other hand part of fluctuations might be anticipated. Both these consideration might induce us a to find a smaller effect size.

To organize our findings, we start by sketching a simple conceptual framework in which agents trade off the monetary gains of cheating against an additively separable moral cost. The basic intuition is that a loss averse agent derives greater marginal utility from the monetary gains of cheating in the loss domain than in the gain domain. Such an agent is hence more likely to cheat when his earnings fall below his reference point.⁴ We follow Koszegi and Rabin (2006, 2007, 2009) theoretical work and treat estimated expectations as the reference point throughout the analysis.

To estimate drivers' expectations, we construct an expected income measure for each driver from a ride-level regression of cumulative income on the number of minutes elapsed in the shift (and its square), the shift start hour, the day of the week, the geographic location of the first ride, and the month of the year. The controls included are readily observable to drivers and are important determinants of earnings in our data. Their inclusion help us to maintain psychological realism while guaranteeing a good fit. We test our main prediction by regressing an indicator for whether a driver commits a rate-code fraud on the gap between the driver's realized and expected performance, controlling for shift invariant characteristics, drop-off time and location of each ride, and weather characteristics. We present evidence from an Income Gap definition that is defined as expected minus realized income (hence, it is positive in the loss domain) and includes the predicted fare obtained in the current ride among the earned income. The results are robust to different Income Gap specifications,

⁴Our model mostly abstracts away from variations in the probability and severity of punishment, though we acknowledge that these are first order determinants of cheating. We also abstract away from another prediction of Prospect Theory, namely that agents will be more risk-seeking in the domain of losses than in gains. The combination of probabilistic punishment and the change in risk posture around the reference point may provide a further force driving more cheating in the loss domain than in the gain domain.

presented in the Appendix B.3.

Endogeneity concerns motivate our instrumental variable strategies. If the decision to cheat is premeditated, then endogenous effort decisions present a potential confound to identification. Indeed, a driver may, at some point during the shift, simultaneously choose to reduce his (unobserved) effort for the next few hours and defraud subsequent customers. The reduced effort is likely to result in a positive Income Gap, inducing us to find a spurious correlation between the driver’s Income Gap and his cheating propensity. To address this threat, we use the Income Gap of all other drivers in the same hour-of-the-day and location as an instrument for a driver’s gap, thereby excluding predetermination concerns.⁵ The results from this approach support the findings of the linear probability model. Unfortunately, this strategy leaves open the possibility of bias arising from the demand side. We attempt to control for these possible influences by including the time of the day and the end location of the ride in all specifications.

The evidence we present could be reconciled by other observationally equivalent mechanisms. If drivers are liquidity constrained for daily needs, even small deviations from expected earnings could cause an increase in the propensity to cheat. This is a credible possibility⁶, which we discuss in section 2.5.4. We have several reasons to believe that liquidity constraints, even if contributing to the effect, are not solely responsible for it. First, we present our main results on ride-level regressions within the shift and identify a high propensity to cheat in the loss domain, even when we restrict the specification to rides in which income

⁵More precisely, as we describe in Section 4, we take the mean of the Income Gap for these other drivers and construct an indicator variable for whether this mean is greater than zero (i.e. in the loss domain). We use this constructed binary variable as an instrument for an indicator for whether the Income Gap of the driver of interest is greater than zero.

⁶On the importance of credit constraints among American households finance see, for instance, Lusardi (2004), Lusardi and Mitchell (2006, 2007). Foley (2011) studies the interaction of liquidity constraints and quasi-hyperbolic discounting in US crime. Specifically, he finds an increase in financially motivated crime over the course of monthly welfare payment cycles.

realizations are within \$25 of the income target. Second, we present regressions that control for cumulative income. When this variable captures the difference between expected daily earnings and needs, we still find an effect on the ride-level within-shift earnings deviation on frauds. Third, our shift level regressions are robust to the inclusions of driver-shift fixed effects. If drivers needs are constant and increase the average propensity to cheat at the same rate over the shift, then driver-shift fixed effects would absorb the effect of liquidity constraints.

A second possibility, closer to our conceptual framework, is that drivers’ moral standards could be a function of deviations from expectations. Our results could then be reconciled by temporary reductions in moral costs rather than loss aversion and reference-dependence.⁷ Ultimately, our data do not allow to separate these possibly complementary explanations.

While field evidence is scant, Loewenstein and Rick (2008) previously posited the role of loss aversion in providing “hypermotivation – a visceral state that leads a person to take actions he or she would normally deem to be unacceptable.” To support this thesis, they point to the goal-setting experiments of Schweitzer et al. (2004). In that paper, students were instructed to solve anagrams and then score their own performance. They find that students are more likely to cheat when provided with exogenously set goals, when meeting the goal is incentivized, and when the subject’s own performance just barely fails to reach the goal. While their experiment does suggest a role for reference-dependence (the “goal” as the reference point), it does not purport to test loss aversion. Since cheating is only relevant for those who failed to reach the goal, they are unable to compare cheating by those in the loss domain against those in the gain domain.

⁷For example, Sharma et al. (2013) use a series of lab experiments to show that when subjects are financially deprived, they are both more likely to cheat for financial gain and to judge less harshly those that cheat for gains.

While other lab experiments in psychology have since further explored this hypothesis ⁸, our paper is the first to explicitly test the role of loss aversion in a natural field context using the observed cheating decisions of agents. Our evidence is complementary to the independent and contemporaneous work on loss aversion and tax payment behavior by Alex Rees-Jones (2013). His paper shows compelling evidence that tax returns display a bunching in the distribution of balances that is consistent with framing of tax payments as losses and tax refunds as gains. The bunching in the distribution is consistent with illegal tax evasion, but could also (as he notes) be reconciled by legal behavior if more effort is exerted to find eligible tax credits and deductions.

Our study also relates to two papers that use field data to study the role of reference-dependence in the context of crime. First, Mas (2006) studies how police performance responds to pay relative to a reference point. He finds that when the police fail to receive an expected wage increase, arrest rates and average sentence lengths decrease, while crime reports increase. These results are suggestive of reductions in effort. Our paper goes one step further to look at clear cases of unethical behavior, the active commission of a crime by agents in the loss domain. A second related paper by Card and Dahl (2012) looks at the role of reference-dependence on domestic violence. They find that reports increase when an NFL team unexpectedly loses a match. While they do study the role of loss aversion in a type of crime, they rely on utility variations derived from sporting outcomes. In contrast, we use utility variations derived from the income stream.

Taxi drivers have also been subjects of several studies within economics. Because drivers choose their own hours and experience plausibly exogenous variation in day-to-day earnings, they have figured prominently in a line of literature on reference-dependent labor supply. A

⁸For example, Kern and Chugh (2009) more directly test the hypothesis by using framing manipulations. In all three of their experiments, students in the loss frame self-reported a higher likelihood of committing a hypothetical unethical act than control students.

first paper by Camerer et al. (1997) showed that drivers display a negative elasticity of labor supply to temporary wage changes and suggested that an explanation for this phenomenon could be found in drivers setting daily earning targets. A number of subsequent papers partly questioned (Farber, 2005, 2008) or reinforced these findings (Crawford and Meng, 2008; Doran, 2009) using different econometric methods and data. In addition, three recent papers consider how taxi drivers charge customers. Keniston (2012) applied a structural model to Indian autorickshaw data to examine the relative welfare implications of allowing bargaining versus fixed-fare regimes. Castillo et al. (2013) used a field experiment to disentangle taste-based versus statistical gender discrimination in bargaining transactions in Peru. Finally, the most closely related study to ours is Balafoutas et al. (2012). They ran a field experiment in Greece to study how taxi cheating, in the forms of over-charging or taking long detours, respond to variation in passengers' presumed information. They show that asymmetric information on the shortest routes and local tariff system is an important determinant of cheating. Unlike their paper, we do not study over-driving, which is the more prevalent type of cheating detected in their study. In contrast, we exploit variation in income relative to expectations to study how overcharging responds to a different type of stimulus, while holding constant information asymmetries. Lastly, four other recent papers consider evidence from the New York City taxi market, including studies of moral hazard in leasing contracts (Schneider 2010, Jackson and Schneider 2011), drivers' learning by doing (Haggag, McManus and Paci 2014), and tipping by customers (Haggag and Paci 2013).

The paper proceeds as follows. Section 2.2 provides a conceptual framework for loss aversion in the context of cheating. Section 3.2 presents the data source and discuss the construction of the fraud measure. Section 2.4 presents the empirical strategy. Section 2.5 reports the results and discusses alternative mechanisms. Section 2.6 concludes.

2.2 Conceptual Framework

We develop a simple conceptual framework to describe the influence of loss aversion on the daily decision to commit an offense. We split the decision problem of each driver in two stages. In the first stage, drivers decide whether to select into cheating based on implicit and explicit costs, permanent income considerations, and severity and probability of punishment, as in the canonical treatment due to Becker (1968).⁹ In the second stage, the driver decides when to cheat on a daily basis. In modeling this second stage, we thus abstract from the risk of punishment, and instead focus on the changes in motivation generated by loss aversion associated with daily income variation. We consider a taxi driver with cumulative income I_t , who decides whether to cheat or not on the next ride on a given day, bracketing each day separately. That is, drivers focus only on daily income fluctuations, rather than integrating these outcomes and expectations over longer horizons.¹⁰ A driver cheats by increasing the fare f to δf , with $\delta > 1$, and he then bears a moral cost of cheating m , where m is an additively separable moral cost.¹¹ Hence, the driver cheats when the gain in utility is greater than the moral cost:

$$\Delta U(\eta, \lambda) = U(I_t + \delta f) - U(I_t + f) > m \quad (2.1)$$

In our formulation, we consider the case of a reference dependent taxi driver with the following utility function:

$$U(x) = (1 - \eta)u(I_t + x) + \eta V(u(I_t + x) \mid u(I_t^r)) \quad (2.2)$$

⁹Unfortunately, our data do not provide us with these variables, and so we do not focus on the first stage in this paper.

¹⁰Narrow bracketing also provides a rationale for ignoring the risk of punishment, as these costs are unlikely to be borne by the driver within the same day as the fraud itself.

¹¹In the empirical part, we control for a vector of time of the day and geographic fixed effects, as well as shift fixed effects.

where η is a weight, u is consumption utility ($u' > 0$, $u'' < 0$), I_t^r is the reference point for daily income and $V()$ is a value function applied to consumption utility¹², as described by Kahneman and Tversky (1979) and formalized by Bowman, Minehart and Rabin (1999). We next show three separate implications of the preference structure depending on the relative position of the reference point I_t^r with respect to the prospects of cheating or not ($I_t + f$, $I_t + \delta f$ respectively), and discuss the predictions for the case of a linear value function¹³:

$$V(\Delta x) = \begin{cases} \Delta x, & \text{if } \Delta x \geq 0. \\ \lambda \Delta x, & \text{if } \Delta x < 0. \end{cases} \quad (2.3)$$

where $\lambda \geq 1$ is the degree of loss aversion and Δx represents the difference between the utility of the actual consumption bundle and the utility of the reference point.

Specifying the condition (2.1) for the reference-dependent driver in (2.2), we obtain with simple algebra that a driver cheats if:

$$\begin{aligned} \Delta U(\eta, \lambda) = & u(I_t + \delta f) - u(I_t + f) + \\ & \eta(\lambda - 1)[u(I_t^r) - u(I_t + f)] * \mathbb{1}[I_t + f < I_t^r < I_t + \delta f] + \\ & \eta(\lambda - 1)[u(I_t + \delta f) - u(I_t + f)] * \mathbb{1}[I_t + \delta f < I_t^r] \\ & > m \end{aligned} \quad (2.4)$$

¹²We follow Koszegi and Rabin,(2006, 2007, 2009) in formalizing the value function as applying to the difference in *utility* rather than physical quantities.

¹³The introduction of diminishing sensitivity yields prediction that depend on the parametrization of the curvatures of both the value function and the consumption utility function u . Therefore, we only present results for the linear value function.

The first line in (2.4) shows that a driver decides to cheat if the increase (recall that $\delta > 1$) in *consumption* utility is higher than the moral cost. For a reference-dependent driver whose daily income is above the reference point, loss-aversion considerations do not affect his motivation to cheat. The second and third lines, both positive terms, show that the benefit from cheating is increased by gain-loss utility considerations, as shown by the indicator functions. This extra benefit is proportional to the difference in utility between cheating and not. Following Loewenstein and Rick (2008), we refer to this increased benefit to cheat as hyper-motivation. The lines distinguish two cases depending on whether committing the crime is sufficient to move the driver above the reference point or just bring him closer to it. In the former case, presented in the second line, loss aversion provides hyper-motivation to cheat only for the gain in utility relative to the reference point, while in the latter case it applies to the full gain in utility between cheating and not. Because of this consideration, loss aversion does not generate a discontinuity in the motivation to cheat as a driver enters the loss domain, but rather a gradual increase. We summarize these observations in the following proposition:

PREDICTION: Controlling for income, a driver has hyper-motivation to cheat when he is in the domain of losses or when cheating brings him above the reference point. The increase is proportional to his degree of loss aversion, given the importance of the gain-loss component in his utility function. For a reference-dependent driver whose daily income is above the reference point, loss-aversion considerations do not affect his motivation to cheat.

Our analysis assumes away stochastic punishment. The introduction of uncertainty in punishment would affect the results in two ways. On the one hand, loss aversion induces first order risk aversion, thereby reducing risk taking. On the other hand, if we were not to restrict to a piece-wise linear value function, diminishing sensitivity would propel risk seeking

in the loss domain. Formalizing risk would require us to formulate further assumptions on how drivers daily bracket the impact of large punishments such as license revocation. We refrain from that and note that the main prediction is robust to the introduction of small risks.

To operationalize the model prediction for our field evidence, we assume that consumption utility is linear over the day, $u(x) = x$, and rewrite the predictions in terms of the Income Gap $G_t = I_t^r - (I_t + f)$. We assume that moral costs are distributed as $m \sim F(m)$. From (2.4), we see then that when $G_t < 0$ the probability of cheating is $P_0 = Pr(c > m) = F(c)$, where $c = \delta f - f > 0$. For the case identified in the second line of (2.4), where $c \geq G_t \geq 0$, the probability of cheating is increased to $P_2 = Pr(c + \eta(\lambda - 1)G_t > m) = F(c + \eta(\lambda - 1)G_t)$. The increase in probability from the baseline P_0 can be written as:

$$P_2 - P_0 = F(c + \eta(\lambda - 1)G_t) - F(c) = \int_c^{\eta(\lambda - 1)G_t} F(s)ds \quad (2.5)$$

And hence the increase in the probability of cheating over this range is a non-decreasing function of the Income Gap. Finally, for the case identified in the third line of (2.4), where $G_t > c$, the probability of cheating is $P_1 = Pr(c + \eta(\lambda - 1)c > m) = F(c + \eta(\lambda - 1)c)$. Thus, for this latter case, the increase in probability of cheating does not depend on the Income Gap. In practice, since c is a small number (mean: 5.26; standard deviation: 6.86), we identify an increase in the probability of cheating using an indicator for $G_t \geq 0$ in the empirical part.

2.3 Data Description and Variable Construction

The data for this study is provided by the Taxi and Limousine Commission (TLC) of New York City. We previously used this dataset to study the effects of defaults on taxi passenger tipping decisions (Haggag and Paci, 2013). Since 2008, all New York City taxi cabs have been endowed with systems that save the GPS coordinates of all pick-up and drop-off locations, as well as ride characteristics (including fares). Our data spans the entirety of 2009, covering the universe of taxi transactions for all licensed NYC taxi drivers in that year. The data contains the following variables: anonymized driver identifier, car identifier, credit card machine company, payment type, ride duration, ride distance, number of passengers, fare, surcharge, MTA tax, toll amount, and pick-up and drop-off latitude, longitude, and time.

During 2009, there were 13,237 taxi licenses, called medallions, operating in NYC. Drivers can be driver-owners, who own both the taxi and the medallion, or can lease their cabs from a company. As described in the Schaller Consulting’s NYC Taxicab Factbook (2006), drivers can lease daily for a single shift (typically an interval of 12 hours or less), weekly, or sometimes for longer periods (usually splitting the cab with another driver for a second shift). In our data we cannot distinguish among these different working arrangements. However, while drivers are subject to different upfront costs depending on their driving arrangements, they always keep all the fare amount and tips earned during a shift and are hence residual claimants of their earnings within a shift.

Our data does not provide the start and end of shifts. Following Farber (2005), we define a shift as a succession of rides such that the time between a ride’s drop-off time and the subsequent pick-up time rides is less than five hours. We further select only shifts that are longer than 30 minutes (we drop 36,986 shifts; 0.49% of the shifts in full sample), shorter than 20 hours (74,671; 0.99%), and shifts where multiple cars are observed (1,298,409; 0.77%).

A final institutional feature is a precondition for our analysis. In New York City customers cannot search for cabs through phone calls or other devices or in any other prearranged fashion, and drivers are prohibited by the regulation to reject customers once they stop for them.¹⁴ In summary, the characteristics of the market matching process and contracts ensures that drivers' earnings are subject to a random component.

The raw data provided by the TLC included 41,256 unique driver identifiers; however, we took a number of steps to clean this data, as detailed in Appendix B.1 (e.g. we removed the 5,189 identifiers associated with fewer than 100 rides). Using our first fraud definition, discussed below, we identified 671,302 episodes of fraud (0.41% of the rides) in the full data. To partially address the issue that some drivers may have unintentionally pressed the incorrect rate code (Rate Code 4), we limit our sample further to drivers that appear to have cheated at least 10 times in 2009 – this sample restriction also attempts to limit our analysis to drivers who have *selected* into the market for cheating (12451 drivers). While we do have data on tips for credit card customers, we do not have this information for cash paying customers. Therefore, following Farber (2008) and Crawford and Meng (2011), we only use fare income to compute expected and realized income. We elaborate on these constructions in Section 4.

To classify fraud, we first reverse-geocoded the data to the Census Tract level. All rides that originate and end in New York City (excluding JFK) should have fares that correspond to Rate Code 1. Under this pricing schedule, the fare accrues as 40 cents for each 0.2 miles while the car is driving over 12 mph and 40 cents for each minute elapsed while the car is

¹⁴As reported by the TLC on the Press Release February 24, 2011, during the period under study, customer refusal punishment was \$200-\$350 for a first offense, \$350-\$500 and a possible 30-day license suspension for a second offense, and a mandatory license revocation for a third offense. Despite the steep penalties, refusal complaints made to the TLC rose from 1,963 in December 2009 to 2,341 in December 2010.

driving under 12 m.p.h.¹⁵ At any given point in time, the meter only accrues one of the two types of charges. Thus, the maximum legal fare for a ride that stays within New York City should be less than the sum of $0.4 \times \text{distance}$ and $0.4 \times \text{duration}$. On the universe of data for 2009, our first pass produces 1,438,625 cases for which the fraud variable takes on value one in the full sample. We then set fraud equal to zero for all rides with either zero recorded duration or distance, as well as all rides with \$45 (JFK-to-Manhattan flat fare) or originating at JFK.¹⁶ These further conditions reduce the number of fraud instances down to 671,302 in the full sample.

The TLC first reported on March 12, 2010 its discovery that "35,558 drivers" had "illegally charged at least one passenger" over a 26-month period – totally 1.8 million instance of rate code fraud. However, this number was later revised downward in a May 14 press release to just 21,819 drivers for a total of 286,000 instances of fraud. Our main fraud identification procedure corresponds to the "Maximum Fare Calculation" outlined in the latter press release and yet generates a higher amounts of frauds; however, we generate a higher number of frauds despite using data covering a slightly shorter time span (the press release corresponded to the period of March 2008 to December 2009). The TLC reported to have used two methods to identify fraud. First, about a third of their data, not provided to us, includes a time-stamping of the Rate Code 4 activation. Thus, the TLC screened out trips where the time-stamp indicated that Rate Code 4 (rather than 1) was activated for less than 20% of total trip time. Indeed, drivers could switch rate code during a ride, pressing the appropriate button in the taximeter, as depicted in Figure 2.1. Second, using our same data, they calculated the maximum possible fare by adding 40 cents per minute (using an assumption that the cab was moving below 12 miles-per-hour for the entire trip) plus 40

¹⁵This tariffs apply to 2009 data. In 2012, the TLC approved an increase in unit fares.

¹⁶Appendix B.2 describes alternative fraud definitions.

cents per 1/5 mile travelled (assuming that the cab was moving above 12 miles-per-hour for the entire trip), as we do. Despite using a similar routine to identify fraud as the TLC, our definition ends up producing more cases than the TLC final report.

Figure 2.1: An example of Rate Code for activation



Source: “TLC Finds Cabbies Overcharged By \$1.1 Million”, NBC New York, May 14, 2010. Accessed online at: <http://www.nbcnewyork.com/news/local/Million-Dollars-93789304.html>.

Our estimates, constructed with an analogous algorithm detect higher amounts of violations. A first reason for the discrepancy could be due to errors in the taximeters recording fare, distance, or duration, for rides on which the TLC had access to the Rate Code variable. Because the fare is computed using distance and duration units, an error in the software could generate a larger fare. If this computational error is present on every ride and is identical on each ride, our results will not be affected (for all the regressions that control for driver-shift fixed effects, since drivers do not change cars during a shift). A second source of measurement error is due to the use of GPS and electronic data. Since Manhattan is populated by a large number of tall buildings, satellite data reporting might be of lower quality as signal refraction might occurs. This factor is likely to be a source of measurement error in our estimates, possibly attenuating our OLS estimates. Finally, the TLC could have

used further criteria for the detection of frauds, which are not known to us.

2.4 Empirical Strategy

To motivate our main analysis, we start in this section by presenting results from shift-level regressions. For each shift, we compute the total income and the proportions of rides on which the drivers cheated. We show that drivers cheated more on low income days, and that the relationship between earned income and the proportion of frauds committed in a shift is non-monotonic. In shifts that displayed unexpectedly low levels of income, drivers cheated even more than predicted by cumulative income and hours alone. We then describe the main ingredients of our within-shift, ride-level analysis. In particular, we detail the construction of the Income Gap measures and the identification strategy used in the estimation. Finally, we discuss possible endogeneity issues and the instrumental variables strategies we employ to address them.

2.4.1 Shift-Level Results

Table 2.1 presents a shift-level analysis of the relationship between income variations and the propensity to commit fraud. The dependent variable in each specification is the number of rides on which we record an instance of fraud, divided by the total number of rides within that shift. Each regression includes drivers fixed effects and uses one observation per driver-shift. Columns 1-4 present results for the full sample of drivers for whom we record at least 10 instances frauds over the course of 2009. Columns 5-8 report results estimated on the restricted sample of shifts for which we record at least one instance of fraud within the shift. This selected sample may be of particular interest, as it abstracts away from the decision of whether to cheat at all on a particular day.

We start in Columns 1 and 5 by showing how the propensity to commit fraud varies with the logarithms of cumulative shift income and shift length (in hours).¹⁷ We find relatively large elasticities in both samples – in the unrestricted sample, a 10% increase in income is associated with a 0.28 percentage point reduction in the fraud proportion (i.e. a $\sim 30\%$ reduction relative to the mean of 0.91 percent). In the remaining columns of the table, we add indicators for whether the shift fare income exceeded various measures of income expectations. These specifications more directly test whether deviations from expectations have an additional effect on the fraud propensity, while controlling for the level of income. Columns 2 and 6 use one of the daily income targets constructed in the reference-dependent labor supply analysis of Crawford and Meng (2011). Among their targets, we present results for their most detailed one, which proxies the target by driver/day-of-the-week-specific sample averages of cumulative fare income up to, but not including, the current shift. The indicator ($\mathbb{1}(Gap_{shift} \geq 0)[CM]$) takes value 1 if the income at the end of the shift exceeded the shift-level proxied expectation. Columns 3,4,7, and 8 use measures that are functions of the ride-level expectations described in Section 4.2. The indicator in columns 3 and 7 ($\mathbb{1}(Gap_{shift} \geq 0)[Mean]$) takes value 1 if the average Income Gap experienced on each of the rides over a shift is positive. The average gap is a noisy indicator that the driver is performing below expectations during a shift. Finally, Columns 4 and 8 present an indicator ($\mathbb{1}(Gap_{shift} \geq 0)[Max]$) that takes value 1 if the maximum Income Gap experienced over the rides of a shift is positive.

Since all measures use imperfect proxies of the underlying shift-level expectations, we expect some attenuation bias. However, our results are indicative of a non-monotonic relationship between fraud propensity and shift income. In all regressions, the coefficient on

¹⁷The income measures exclude fraud-derived income. We impute the non-fraudulent fare from a linear prediction on distance and duration and their squared on the full sample.

log income is negative and statistically significant. Each of the proxies suggest an additional effect of being below expectations relative to the unconditional mean. Our preferred shift level indicator, $(1(Gap_{shift} \geq 0)[CM])$, constructed following Crawford and Meng (2011), suggests an increase ranging from 0.9% (column 5) to 8.9% (column 2). The *Mean* and *Max* indicators, constructed from the average and the maximum Income Gap experienced by a driver on all of the rides within the shift, are a more noisy measure of performance with respect to expectation. The *Mean* indicator picks up a higher propensity to cheat when on average, the driver operates below income during the shift. The *Max* indicator takes value 1 if the maximum Income Gap over all the rides is positive. For this indicator the relationship is attenuated. This is not surprising since in most shifts drivers are at some point lagging below expectation. Yet, in the restricted sample (col. 6), we still find an effect, albeit a marginally significant one.

Table 2.1: OLS: Proportion of Rides with Fraud, by Shift

| | Unrestricted Sample | | | | Restricted Sample (Fraud in Every Shift) | | | |
|--|------------------------|------------------------|------------------------|------------------------|--|------------------------|------------------------|------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\mathbb{1}(Gap_{shift} \geq 0)[CM]$ | | 0.0008*** (0.0001) | | | | 0.0013** (0.0007) | | |
| $\mathbb{1}(Gap_{shift} \geq 0)[Mean]$ | | | 0.0004** (0.0001) | | | | 0.0014* (0.0007) | |
| $\mathbb{1}(Gap_{shift} \geq 0)[Max]$ | | | | 0.0001 (0.0002) | | | | 0.0025* (0.0014) |
| Log(Shift Income) | -0.0284*** (0.0019) | -0.0270*** (0.0018) | -0.0293*** (0.0024) | -0.0295*** (0.0023) | -0.1880*** (0.0054) | -0.1845*** (0.0055) | -0.1886*** (0.0065) | -0.1895*** (0.0063) |
| Log(Shift Hours) | 0.0244*** (0.0016) | 0.0240*** (0.0016) | 0.0253*** (0.0020) | 0.0254*** (0.0020) | 0.1019*** (0.0056) | 0.1043*** (0.0058) | 0.1033*** (0.0066) | 0.1037*** (0.0066) |
| Observations | 3,075,205 | 2,990,211 | 2,061,046 | 2,061,046 | 205,412 | 192,357 | 141,096 | 141,096 |
| r ² | 0.5840 | 0.5923 | 0.5997 | 0.5997 | 0.8200 | 0.8209 | 0.8210 | 0.8210 |
| Mean of Dep. Variable | 0.0091 | 0.0089 | 0.0095 | 0.0095 | 0.1361 | 0.1388 | 0.1391 | 0.1391 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

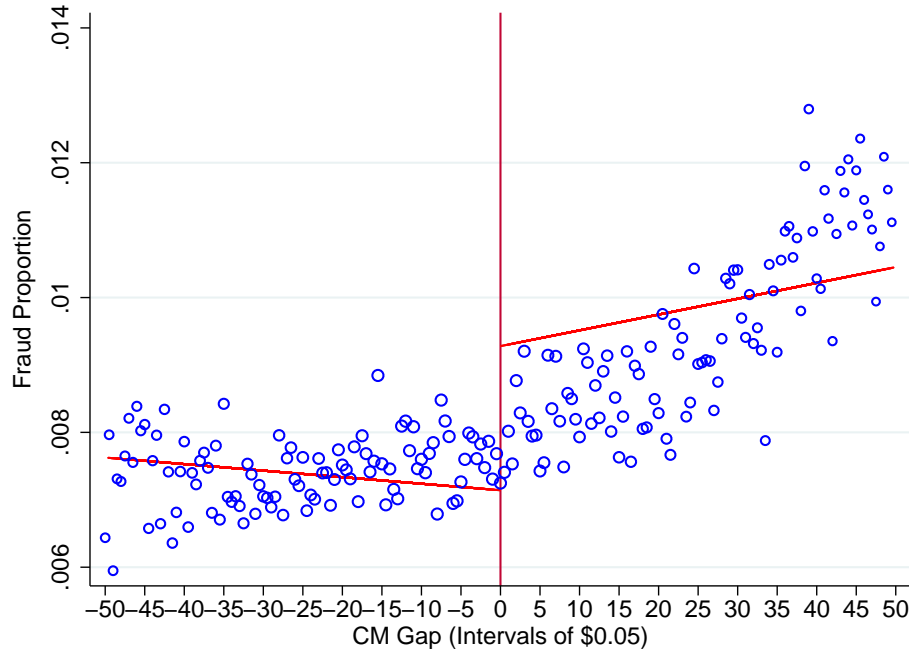
Notes: Robust standard errors clustered at the driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day. The Crawford and Meng Proxy is estimated over a smaller sample since, as explained in the text, the first shift for each driver/day-of-the-week is excluded from the analysis.

Figure 2.2 demonstrates the relationship between fraud proportion during a shift and income, using our preferred measure, introduced by Crawford and Meng (2011). The graph restricts to the interval of \$50 around zero. Observations in the positive domain of the graph are those shifts for which the total realized shift income falls below the proxied expectation (i.e. the average of shift incomes over all previous shifts on that same day of the week for that driver).¹⁸ That is, as with the regressions, positive values correspond to the loss domain. On the y-axis we report the proportion of frauds within the shift. The dots report

¹⁸As noted above, both of these measures exclude fraudulently derived income.

the average for each \$0.50 bin on either side of zero, and are weighted by the number of observations. The regression lines are constructed using the full sample, i.e. including shifts that ended with an income beyond \$50 above or below the estimated target. The figure shows a higher proportion of frauds when drivers' income is below target, consistent with the regression in Column 2 of Table 2.1 that additionally controlled for (log) income and hours. The regression lines show that the positive and negative realizations have different means. For positive Income Gap, we observe a slightly positive slope. As the measure of performance is noisy, this would be consistent with the targets not precisely picking up the actual expectation of the drivers, as well as with drivers not attending to small deviations. Indeed, we see that the average fraud is markedly higher when the deviation is larger than about \$25, when it is likely more noticeable that the shift's income is lower than expected.

Figure 2.2: Crawford and Meng's Distance from Expectation Measures: Proportion of Rides with Fraud, by Shift



Notes: Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature $> 80F$, temperature $< 30F$), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day. CM Gap is the Crawford and Meng Proxy estimated as deviation of shift income from the previous driver/day-of-the-week shift average from the beginning of the sample. Positive deviations correspond to worse income realizations (loss domain).

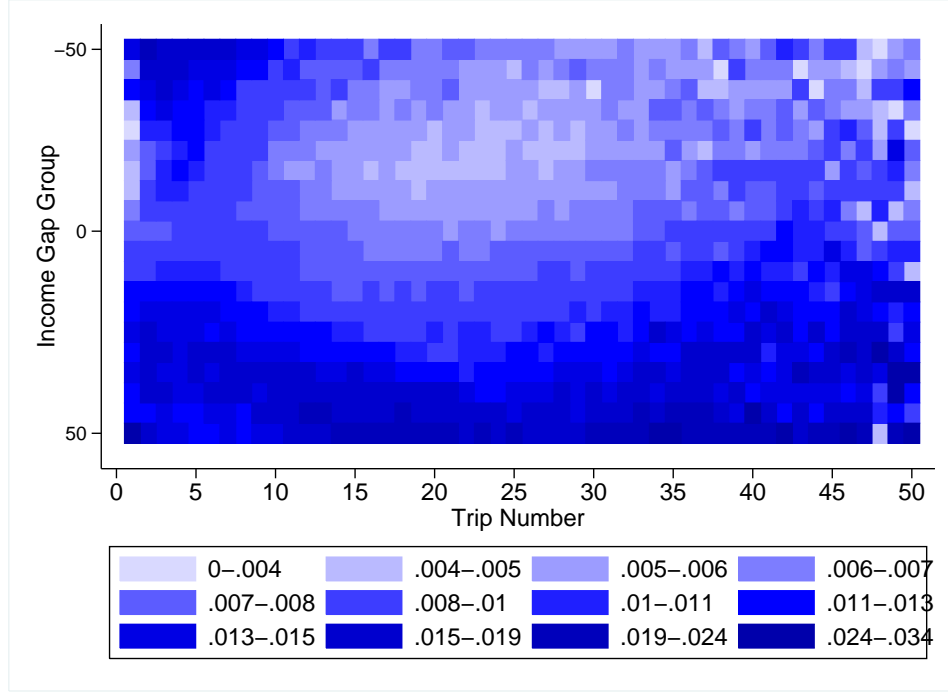
In these within-driver, across shifts comparisons, the relationship between fraud proportion and deviations from the shift average income performance, as measured by the three indicators reported, suggests that loss aversion might play a role in cheating behavior. However, the desire to smooth out unexpected shocks in shift income could be reconciled with consumption smoothing. If drivers do not have access to a well functioning credit market, income smoothing might be a precondition for consumption smoothing. This is a possibility that cannot be ruled out when looking at shift-level regressions. Moreover, the decision of when to quit a shift is endogenous and depends on income. For example, if drivers choose

their fraud amount using a constant fraud-per-day rule of thumb (e.g. 2 per day) and quit earlier on shifts with unexpectedly low income realizations (though this would be inconsistent with Crawford and Meng, 2011), then we would find higher fraud proportions on below expectation shifts – a spurious relationship relative to our theory.

To move beyond these issues, the rest of our analysis focuses on small-sized fluctuations in income within a shift. Although it is possible to imagine that tightly binding liquidity constraints operate in a similar way at the daily level, concerns about picking up smoothing motives are attenuated. Moreover, we will be able to deal with drivers' effort decisions with an instrumental variable approach that takes advantage of demand fluctuations at the hour-location level.

Figure 2.3 graphically illustrates the relationship between frauds and trip number. It vertically divides the Income Gap described in 2.2 into twenty intervals ranging from an Income Gap of \$50 to -\$50, on the y-axis. On the x-axis, it illustrates the trip number. The colored panels represent the average fraud in each bin. Moving from lower to higher bins, we observe a decrease in the empirical probability of fraud. The graph demonstrates that fraud episodes are more prevalent when the Income Gap is larger, that is when drivers are in the domain of losses. The figure also shows that frauds are markedly more likely to occur in the domain of losses for trip numbers in the range 10 to 25. For the first rides, the proportion of frauds seems to be unaffected by the level of the Income Gap. We also observe that, over 10-25 range rides, the increase in fraud propensity gains momentum when the difference from expectations is more marked (i.e. when the driver lags behind income by at least \$20).

Figure 2.3: Average Fraud by Income Gap measure and Trip Number within a Shift



Notes: Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare.

2.4.2 Construction of the Income Gap Measures

To test our prediction that drivers below expected income are more likely to commit fraud, we estimate a ride-level income expectation for each driver within our sample. In particular, we estimate a linear function $\kappa_i(t(r), \xi_s)$ that indicates the amount a driver expects to have earned from the beginning of the shift up to time $t(r)$, the pick-up time for ride r , controlling for shift fixed characteristics ξ_s . For driver i who has worked for $t(r)$ minutes (e.g. 200 minutes) since the beginning of his shift, the function's value depends on the time of day (night/day), the day of the week (e.g. Thursday), the month (e.g. October), and the broad location of the first pick-up of the shift (e.g. Uptown Manhattan). The fare for each

ride is composed of two parts: a \$2.50 flat fee and a component that is a function of the distance and duration of the trip. Thus, every extra ride is associated with a sure payment, corresponding to the fixed component, and a stochastic component which, together with search time between rides, is the source of our identifying variation.¹⁹ Starting with the measure of cumulative income earned by each driver i during shift s up to ride r , $I_{i,s,r}$, we difference out the fixed component of each fare, obtaining the term $Z_{i,s,r} = I_{i,s,r} - 2.5 * r$. We then run the following regression separately for each driver:

$$Z_{i,s,r} = \kappa_i(t(r), \xi_s) = \beta_1 t(r) + \beta_2 t^2(r) + n_{i,s} + d_{i,s} + m_{i,s} + g_{i,s} + u_{i,s,r} \quad (2.6)$$

Where $t(r)$ is minutes elapsed since the pick-up time of the first ride of the shift, $t^2(r)$ its square, and $n_{i,s}$, $d_{i,s}$, $m_{i,s}$, $g_{i,s}$ are, respectively, fixed effects for daily (4am-4pm) shifts, day of the week, month of the year, and geographic location of the first ride of the shift. The fixed effects are included because incomes tend to vary across nights and daily shifts, by day of the week and over the year.²⁰ Moreover, the geographic location of the first shift is partly chosen by the driver (e.g. a driver can decide to start the shift at JFK airport).²¹ All these factors are possibly anticipated by the driver and we assume that they are taken into account when forming the income expectation.

We use the predicted values from the regression to construct a measure of the expected

¹⁹Drivers are prevented by the law from screening customers based on their destination. However, it is possible that on some transactions cherry picking of customers still occurs. The inclusion of time and geographic fixed effects in our ride level regressions aims at addressing this concern.

²⁰Appendix B.3 reports results from this decomposition.

²¹We divide the New York City area in the following groups: The Bronx, Staten Island, Brooklyn, Queens - excluding airports, Manhattan-Downtown, Manhattan-Midtown, Manhattan-Upper East, Manhattan-Upper West, Manhattan - Uptown, JFK airport, La Guardia airport.

cumulative income for each driver in shift s and ride r , $\hat{I}_{i,s,r}(t(r), \xi_s) = \hat{Z}_{i,s,r} + 2.5 * r$. The expectation seeks to capture the opportunity costs of time spent working in a shift from the driver's perspective. The inclusion of time rather than ride number in the regression captures the profitability of a shift – being at the second ride after 200 minutes differs from being at the fifth ride in that the first situation suggests a slow day. The psychological reaction we posit is one that compares the income over time without fully adjusting to the contingency of the specific shift. The Income Gap captures this discrepancy between the expected and the realized return from the time spent in the activity, adjusting for a handful of observable shift level characteristics (e.g. the returns on a Thursday versus a Sunday).

We construct the Income Gap measure as $Gap(i, s, r) = \hat{I}_{i,s,r} - (I_{i,s,r-1} + \hat{f}_r)$, where \hat{f}_r is the predicted fare on ride r .²² The driver is assumed to compare the expected earnings usually obtained in a shift up to and including ride r , to the income acquired in the current shift by the end of the ride that just started. Thus, the driver takes into account the financial consequences of the current ride on his Income Gap.

Our results are robust to other measures of expectations. First, we constructed a measure of expectations that neglects the (predicted) impact of the current ride on cumulative income. In this measure, a driver compares the amount of money usually earned, $\hat{I}_{i,s,r-1}$, with the amount of money earned on the current shift up to the previous ride, $I_{i,s,r-1}$. Second, we constructed measures of the Income Gap that include the income earned on cheated rides, which might very well be the most salient measure of cumulated income for the drivers. Third, we computed our main Income Gap measure with different polynomials over the minutes elapsed since the pick-up time of the first ride of the shift. While our main measure includes a second order polynomial in this term, our results are robust to different choices,

²²We estimate the current ride's expected fare based on the destination location, duration and distance on the full sample of drivers to reduce noise.

such as third or fourth order polynomials. We present detailed results in Appendix B.3.

Our analysis rests on there being meaningful exogenous variation in income within a given shift. We exclude fraud-derived income from our measures of expectations and income realizations since fraud derived income can induce a mechanical relationship consistent with our theory. For example, consider a driver who earns exactly as expected up to ride r and cheats on ride r for reasons unrelated to the gap. If we use realized income and include the value of the fraud, then a driver may be above expectation for rides $r + 1$ onward (simply because the fraud-derived income puts him above expectation). Thus, if we compare propensity to cheat at ride level using shift level fixed effect, the gain position from ride $r + 1$ puts the driver *relatively* in the loss domain for ride r , inducing us to find an effect.²³ To exclude fraud from income, we construct a counterfactual fare from a regression of fare on distance, duration and their squares on rides in which no fraud is observed ($R^2 = 0.974$) and impute this value for the fare in all instances of fraud. We present results for measures of the Income Gap that include frauds derived income in Appendix B.3, and show that they are qualitatively similar.

Our results rests on the accuracy of the Income Gap regression to measure the distance between realized and expected earnings. However, in order to argue that preferences display reference-dependence to expectations, a researcher should be able to manipulate expectations while holding constant earnings. We are currently working on this exercise in a complementary experimental work.

²³Using a measure of income that excludes frauds has also its drawbacks. Primarily, it does not correspond to what drivers can see on their taximeters. Moreover, this approach is problematic in presence of positive autocorrelation in fraud. If a driver who cheats at ride r is above income on subsequent rides in terms of income including frauds, but not in terms of our measure of income, and cheats more for positive autocorrelation (e.g. because moral costs are temporarily reduced after a first fraud), then we would associate frauds with being below expected income.

2.4.3 Identification and Instrumental Variable Analysis

To test our main prediction that drivers are more likely to overcharge customers when their income falls below their expectations, we present ride-level regressions of the probability of committing a fraud against an indicator for whether the Income Gap is positive (i.e. the agent is in the loss domain). Our conceptual framework predicts a higher fraud propensity for a positive Income Gap. We present specifications that include an indicator for whether the gap is positive, abstracting from the narrow region of small positive Income Gap over which the propensity to cheat is predicted to be a non-decreasing function of the Gap. In all of our regressions, we control for other observable determinants of fraud, including ending location of the ride, hour of the day, and day of the week fixed effects (Appendix B.3 reports the contributions of these fixed effects to cheating probability). Moreover, drivers might find it more risky to cheat from the beginning on longer and more expensive rides, where the difference in amount might be more noticeable to the costumers, so we also include distance and duration of the ride (which drivers can form a prediction over when the customer enters the taxi and requests their drop-off destination). These considerations suggest the following baseline specification:

$$F_{i,s,r} = \beta \mathbb{1}(\text{Gap}(i, s, r) \geq 0) + \mathbf{X}\gamma_{i,s,r} + \xi_{i,s} + \epsilon_{i,s,r} \quad (2.7)$$

Where $F_{i,s,r} = 1$ is an indicator driver i commits fraud on ride r of shift s . The indicator $\mathbb{1}(\text{Gap}(i, s, r) \geq 0)$ takes value one if driver i 's income is below expectation, and \mathbf{X} contains ride characteristics (distance and duration) and dummies for hour of day (1-24), day of the week (0-6), drop-off location²⁴, and dummies for whether the hourly minimum temperature is 30F, the maximum temperature is above 80F and rain is detected in either of the New

²⁴We partition NYC using the same classification as in Farber (2005).

York City weather stations. Finally, the vector $\xi_{i,s}$ is a driver-shift fixed effect.

We further seek to isolate the effects of the income variation in a particular shift from other unobserved driver-day specific determinants of fraud by including three types of fixed effects. We present specifications that control for time invariant characteristics of the driver-shift, hence capturing average changes in propensity to cheat related to unobservable that vary over longer time spans (e.g. rent payments). Second, we expand the fixed effects to the driver-level, in order to further exploit the between-shift variation in addition to the within-shift variation. Finally, we present results for driver-day-of-the-week-hour fixed effects in order to capture average determinants of the Income Gap that have to do with driver habits (e.g. the driver can take a lunch break at noon on Thursday). We estimate a Linear Probability Model and a Conditional Logit model, both for the full sample and restricting to rides for which the Income Gap is in the $\pm \$25$ interval around zero. For such small deviations, liquidity constraints are likely to be less powerful motivators of cheating behavior.

There are a number of potential confounding factors that lead us to use an instrumental variable approach for the Income Gap in the context of our LPM. A first source of endogeneity arises if drivers, at some point during the shift, jointly decide to commit fraud in subsequent rides and to work less intensively (due to anticipating higher, fraud-inflated, income in later rides). For this case of a predetermined fraud choice, our observational evidence could be explained by reduced search effort, as this will be likely reflected in a higher Income Gap. To deal with this possibility, we exploit aggregate demand fluctuations. For each driver we instrument the Income Gap measure with the average gap experienced in the same hour by all the other drivers in the sample who are picking up customers in the same geographic location (we use the same partition of the city described in section 2.4.2). This allows us to capture market conditions that are independent from driver's predetermined decisions.

A drawback of this identification strategy is that demand composition might itself be a source of bias. For instance, drivers may lag behind expected income in moments when there is a higher proportion of customers who are uninformed about the fare structure (e.g. tourists). Our specification seeks to address violation of the exclusion restriction that spurs from asymmetric information by adding hour and place fixed effects for each ride.

Finally, we note that the Income Gap is a *generated regressor* (Pagan, 1984). The usual assumptions for estimation are sufficient for consistency in the linear probability model, however the standard errors ignore the sampling variation in the regressor, which is generated using the same sample. The use of an instrument that satisfies the exclusion restriction eliminates this problem with inference. Hence, we do not present adjusted standard errors in the LPM section, which would be computationally very intensive as every driver in the sample has a specific expectation function.

2.5 Results

2.5.1 Summary Statistics

Table 2.2 presents summary statistics on our primary regression sample. This sample is restricted to rides on which it is possible to detect fraud, reducing the sample from 73,073,160 to 69,625,780 observations.²⁵ The first column presents mean and standard deviations on the first ride, while columns 2-4 span the 5th, 20th, and last ride (which may come before or after the 20th ride on different shifts), respectively. Finally, the last column presents results

²⁵Specifically, these are rides which satisfy the following conditions: (1) both started and ended within New York City, (2) neither started nor ended at the census tract associated with John F. Kennedy Airport, (3) the payment type was not “No Charge”, (4) neither ride duration (computed using the difference between start time and end time) nor distance are equal to zero, (5) did not have a fare equal to \$45 (the Kennedy Airport flat fare).

for the full sample. The penultimate row of Panel A presents the predicted fare used to construct our measure of income that excludes frauds. The average for *Predicted Fare* is not the same as that of *Fare* since the prediction is constructed over the full sample of rides, while here we restrict to our regression sample, which includes only rides over which fraud is possible. The probability of fraud is around 1%, and has a higher mean on the first and last rides of the shift than the 5th or 20th rides. As in Figure 2.2, we see that frauds are more likely to happen on the first and last ride of the shift.

Table 2.2: Summary Statistics - Ride level

| | (1) | (2) | (3) | (4) | (5) |
|---|-----------|-----------|-----------|-----------|------------|
| | Trip=1 | Trip=5 | Trip=20 | Last Trip | All |
| Panel A: Ride-Level Variables | | | | | |
| Duration (Minutes) | 13.4548 | 11.4082 | 11.1671 | 13.3397 | 11.3230 |
| | (10.7485) | (7.9970) | (7.3896) | (11.3765) | (7.8225) |
| Distance (Miles) | 3.2686 | 2.2994 | 2.6000 | 3.9217 | 2.5202 |
| | (3.2062) | (2.2336) | (2.4047) | (3.5305) | (2.4027) |
| Fare | 11.0744 | 9.0075 | 9.3886 | 12.0756 | 9.3151 |
| | (7.3615) | (5.3073) | (5.5316) | (7.9415) | (5.5671) |
| Predicted Fare | 10.9928 | 8.9209 | 9.3045 | 11.9660 | 9.2285 |
| | (7.1948) | (5.1717) | (5.4062) | (7.6126) | (5.4363) |
| $Pr(Fraud)$ | 0.0109 | 0.0101 | 0.0076 | 0.0103 | 0.0088 |
| | (0.1039) | (0.0998) | (0.0869) | (0.1011) | (0.0933) |
| Panel B: Income Realizations, Expectations, & Gaps | | | | | |
| Cumulative Income | 10.9961 | 51.3552 | 184.5148 | 233.5395 | 133.8621 |
| | (7.2634) | (22.0021) | (35.6113) | (84.3129) | (88.7124) |
| Expected Income | 14.2662 | 52.4539 | 185.0658 | 233.7426 | 134.8260 |
| | (9.0737) | (22.1111) | (30.5446) | (76.1249) | (85.4284) |
| Gap(i,s,r) | 2.9207 | 1.2445 | 0.7443 | 0.1180 | 0.3818 |
| | (8.2459) | (14.5676) | (22.6645) | (30.0417) | (20.0826) |
| N | 2,726,856 | 2,867,509 | 2,001,056 | 2,820,122 | 69,625,780 |

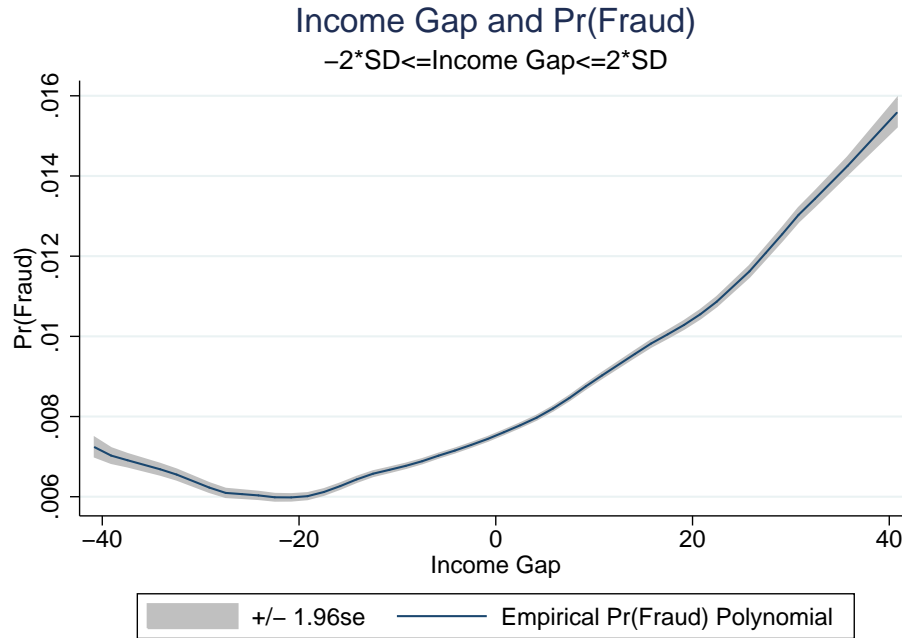
Notes: Standard deviations in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

Panel B of Table 2.2 reports statistics on the income expectation proxies, realizations, and the implied Income Gap. On average, drivers end their shifts with \$234. The second row reports the estimated *Expected Income*, $\hat{I}_{i,s,r}(t(r), \xi_s)$, which is used to construct the Income Gap, $Gap(i, s, r)$, displayed in the third row. As discussed in Section 2.4.2, we construct *Expected Income* using a linear predicted value from a regression of the *Cumulative Income* on several variables. Since these linear predictions produce negative values for 0.4% of the observations (among the negative forecasts, 77.5% were associated with the first ride of the shift), but drivers never hold negative expectations, we drop all negative values. Moreover, the regression is constructed over the full sample, including rides in which fraud is not possible. As a result, the Income Gap deviates from mean zero, and is slightly positive for the early rides of the shift.²⁶ To attenuate concerns arising from this type of misspecification, our linear probability model predictions include ride number fixed effects.

Figure 2.4 graphically depicts the relationship between the empirical probability of fraud and our main income gap measure, for an interval of two standard deviations of the income gap around zero. The probability of fraud starts to increase as the gap become positive. The underlying noise in the data, and the difficulty in capturing the underlying expectation process, generate a smoother transition from the domain of gains to that of losses than our conceptual framework would predict. Moreover, we observe a continuous increase when the gap is positive. If drivers only notice deviations from the gap as they become more sizeable, this steadier increase in the propensity to cheat is still largely consistent with our framework. We comment further on this point in the following section.

²⁶Appendix B.3 reports graphical evidence on the Income Gap.

Figure 2.4: Empirical Fraud Probability by Income Gap measure



Notes: Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare.

2.5.2 Linear Probability Model and Conditional Logit Model Estimates

Table 2.3 reports linear probability model estimates for the specification detailed in equation (2.7). The table shows the sensitivity of our estimates to various assumptions on the appropriate functional form and sample restrictions. Columns 1-3 report results for the full sample, while columns 4-6 report results estimated over rides on which the deviations from the expected income is less than \$25. This latter set of columns ensures that our results are not driven by responses to large shocks (the standard deviation of the Income Gap is slightly above \$20 on the full sample). Panel A reports results from the unrestricted sample, while

Panel B is limited to shifts on which at least one instance of fraud is recorded. Columns 1 and 4 include shift-driver fixed effects, columns 2 and 5 include driver fixed effects, and columns 3 and 6 include driver-day-of-the-week-hour fixed effects.

Our estimates are qualitatively robust to the inclusion of different fixed effects and to the sample restrictions. Column 1 of Panel A shows that drivers are roughly 7% more likely to commit fraud on rides in which they fall below expectations ($\mathbb{1}(Gap(i, s, r) \geq 0) = 1$) relative to rides within that same shift in which they are above expectation (on a mean fraud probability of 0.88%). Columns 2 and 5 show that the effect is even larger (18%) when we only include driver fixed effects, allowing for the additional between-shift variation. Column 3 and 6 suggest that the relationship between the Income Gap and the cheating probability is even larger when we take into account usual patterns of behavior by controlling for driver-hour-day fixed effects, that could be otherwise correlated with our gap measure. This last specification also suggests an increase in the likelihood of cheating by about 18%.

Table 2.3: Linear Probability Model: Fraud Probability, by Ride

| | Full Sample | | | −\$25 ≤ Gap(i,s,r) ≤ \$25 | | |
|--|-----------------------|-----------------------|-----------------------|---------------------------|-----------------------|-----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A: Unrestricted Sample | | | | | | |
| 1(Gap(i,s,r) ≥ 0) | 0.0006*** (0.0000) | 0.0016*** (0.0000) | 0.0016*** (0.0000) | 0.0004*** (0.0000) | 0.0007*** (0.0000) | 0.0006*** (0.0000) |
| Pr(Fraud) | 0.0088 | 0.0088 | 0.0088 | 0.0083 | 0.0083 | 0.0083 |
| Observations | 69,625,780 | 69,625,780 | 69,625,780 | 39,328,345 | 39,328,345 | 39,328,345 |
| r ² | 0.4037 | 0.2441 | 0.2828 | 0.4057 | 0.2560 | 0.2973 |
| Shifts | 3,036,570 | 3,036,570 | 3,036,570 | 2,027,101 | 2,027,101 | 2,027,101 |
| Drivers | 12,451 | 12,451 | 12,451 | 8,451 | 8,451 | 8,451 |
| Panel B: Restricted Sample (Fraud in Every Shift) | | | | | | |
| 1(Gap(i,s,r) ≥ 0) | 0.0029*** (0.0005) | 0.0142*** (0.0004) | 0.0052*** (0.0005) | 0.0005 (0.0005) | 0.0071*** (0.0004) | -0.0005 (0.0005) |
| Pr(Fraud) | 0.1199 | 0.1199 | 0.1199 | 0.1150 | 0.1150 | 0.1150 |
| Observations | 5,096,109 | 5,096,109 | 5,096,109 | 2,843,360 | 2,843,360 | 2,843,360 |
| r ² | 0.3387 | 0.2775 | 0.3899 | 0.3445 | 0.2814 | 0.3863 |
| Shifts | 205,517 | 205,517 | 205,517 | 119,159 | 119,159 | 119,159 |
| Drivers | 12,451 | 12,451 | 12,451 | 7,244 | 7,244 | 7,244 |
| Fixed Effects | Shift-Driver | Driver | Driver-Day-Hour | Shift-Driver | Driver | Driver-Day-Hour |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the fixed effect level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

Consistently with Figure 2, columns 4-6 show that the estimated effect sizes are attenuated when we restrict to an interval for the Income Gap of \$25 around zero. While in the full sample regressions we identify the positive indicator out of the 52% of shifts in which it switches from positive to negative – across fixed effects specifications, in this subsample the indicator is turning from positive to negative, or viceversa, in 78% of the shifts. The

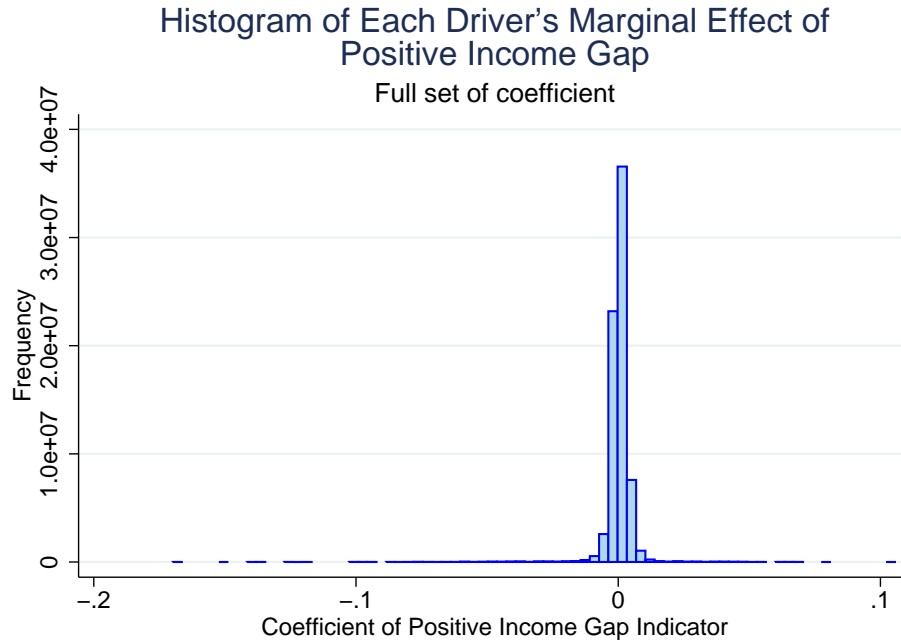
increase in cheating is still sizeable (4.8%-8.4%), suggesting that even small deviations from expectations are sufficient to elicit a behavioral response. However, larger gaps magnify the response. This could be due to drivers not attending to small deviations from expectations. Moreover, it is also possible that liquidity constraints that would trigger a larger response become binding only when deviations become sizeable.

Panel B restricts to shifts in which at least one episode of frauds is detected. If drivers are prevented from cheating on some shifts based on unobservable determinants, for instance if some managers at specific companies are more likely to check the trip reports, our estimates could be biased downward.²⁷ In this restricted sample, our estimates range from 2.4% to 11.8%. Finally, looking at estimates in the $-/+$ \$25 dollars gap sub-sample, we see that estimates are not significant in two of the specifications (cols. 4 and 6). The coefficient remains significant and of similar size in the regression that only includes drivers fixed effects.

In Figure 2.5 we present the distribution of the marginal effect on positive income gap estimated from separate linear probability model regressions for each driver. The mean and mode of the coefficients indicates an increase in the disposition to cheating when the driver falls in the loss domain with respect to expectations. Nonetheless, drivers show heterogeneous responses.

²⁷Daily lease drivers can in principle switch between garages. We do not think differential detection by managers of the taxi company to be a highly likely possibility. The manager should perform a computation similar to our Maximum Fare calculation on every ride of the trip-sheet, an event that seems unlikely to occur. Indeed, we are not aware of any company manager reporting the rate code manipulation to the TLC.

Figure 2.5: Distribution of Marginal Effect on Positive Income Gap, By Driver



Notes: Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day. Coefficients obtained from separate linear probability model regressions for each driver. Positive deviations correspond to worse income realizations (loss domain).

Table 2.4 reports results for a conditional logit model. For this model, the error term $\epsilon_{i,s,r}$ in (2.7) is assumed to be i.i.d. Type I Extreme Value distributed, conditional on the driver-shift (or driver-hour-day) fixed characteristics. The Conditional Logit allows us to get a more precise fit of the Conditional Expectation Function. Because fraud in our data is a low probability event, the non-linear model could produce different marginal effects, as we are close to the tail of the distribution. However, the sample for these results is a restricted one, since the conditional logit drops groups (driver-shifts or driver-hour-day-of-the-week) over which no fraud is detected. We find estimates that are similar to the LPM, but slightly larger in magnitude than the comparable results found in Panel B of Table 2.3. The

quantity $\Delta Pr(Fraud)$ reports the predicted percentage increase in probability of fraud when the driver moves from the gain to loss domain²⁸ – this model predicts an increase ranging from 6.4% to 22.4%, depending on the fixed effects included. For the sample restricted to +/- \$25, we find again a lower response, ranging from 4.5% to 9.6%. Across specifications and sample restrictions, this approach confirms the patterns observed in the linear regressions.²⁹

²⁸To construct the change in cheating probability, we round the coefficient c to the fourth digit and use it to construct the new odds as $n = o * (1 + e^{c-1})$, where o is the baseline odds ($o = \frac{p}{1-p}$, where p is the regression sample frequency of the dependent variable).

²⁹Computational constraints prevent us from computing the Conditional Logit for driver fixed effect specifications.

Table 2.4: Conditional Logit Model: Fraud Probability, by Ride

| | Full Sample | | $-\$25 \leq \text{Gap}(i,s,r) \leq \25 | |
|--|-----------------------|-----------------------|--|-----------------------|
| | (1) | (2) | (3) | (4) |
| $\mathbb{1}(\text{Gap}(i,s,r) \geq 0)$ | 0.0762*** (0.0064) | 0.2122*** (0.0062) | 0.0497*** (0.0069) | 0.0969*** (0.0064) |
| Observations | 5,089,223 | 15,479,212 | 2,604,174 | 7,585,485 |
| Pr(Fraud) | 0.1187 | 0.0392 | 0.1230 | 0.0427 |
| Δ Pr(Fraud) | 6.9 | 22.4 | 4.4 | 9.6 |
| Pseudo r ² | 0.0392 | 0.0854 | 0.0444 | 0.0683 |
| Log Likelihood | -1011823.4 | -1427083.7 | -526206.1 | -745473.2 |
| Drivers | 1,019 | 2,475 | 864 | 2,198 |
| Shifts | 139,872 | 498,988 | 118,924 | 465,687 |
| Fixed Effects | Shift-Driver | Driver-Day-Hour | Shift-Driver | Driver-Day-Hour |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the fixed effect level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature $> 80F$, temperature $< 30F$), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

2.5.3 Instrumental Variable Estimates

Table 2.5 presents results for our instrumental variable strategy. For each day of the year and hour of the day, we compute the average Income Gap for each driver operating in the market (e.g. if a driver has 3 pick-ups in that hour, this variable will take the mean of the Income Gap over those 3 rides). We then split up the city again in the 12 different areas described above, using the same partition used for the fixed effects in the expectation

regression reported in Section 2.4.2. For each driver, we isolate the sample of other drivers that happen to have at least one pick-up in the same location and hour-of-the-day as the ride of interest (i.e. for ride r by driver i during shift s at 9:20am in midtown on October 2nd 2009, we take the average gap of all drivers that have at least one pick-up in midtown between 9am and 10am on October 2nd). Next, within this sample of isolated drivers, we take the average of their hour-of-the-day specific average to construct the *Other Drivers'* Income Gap composite (e.g. the average of hourly gap for that hour-day of the drivers who are operating in that same location). Finally, we construct an indicator variable for whether this composite is positive, and we use this indicator variable as an instrument for whether driver i is above target on ride r in shift s .

The first three columns of Table 2.5 report this regression for shift, shift-driver and driver-day-of-the-week-hour fixed effects, respectively. Columns 4-6 replicate the specification on the restricted sample. The first stage indicate a strong correlation across the fixed effect included and across both samples. We find that the estimated effects are much larger in magnitude than those reported in Tables 2.3 and 2.4. Our preferred specification, which controls for time invariant characteristics over each driver-shift, is reported in Column 1. This specification allows us to look at the effect of aggregate shocks while controlling for unobserved determinants of cheating that do not vary over the shift, such as weekly needs. It reports a coefficient of 0.005, which relative to the unconditional mean of 0.0088, represents a 56.8% increase in the probability of cheating. In the restricted sample, the estimated coefficient of .0151 (column 4) corresponds to a 12.6% increase relative to the mean of 0.1199. Overall, these results suggest that OLS and Logit may largely understate the magnitude of the effect.

Deviations in drivers income that are due to the driver's effort decisions are not expected to trigger an increase in cheating in our framework, as such an effort decision would be real-

istically accompanied by a change in expectation. Our IV strategy identifies our coefficients by isolating variation in the Income Gap that is due to market fluctuations. Restricting to this more likely unexpected dimension of the Income Gap, the size of the effect is magnified. While our methodology does not allow us to identify the degree to which earnings fluctuations are unanticipated, the IV results at least allow us to show that the relationship is large and robust to isolating those gaps that are plausibly independent of endogenous effort decisions.

Table 2.5: IV: 2SLS Estimates and First-Stage: Fraud Probability, by Ride

| | Full Sample | | | Restricted Sample | | | Airport | |
|---------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| IV Estimates - DepVar: Fraud | | | | | | | | |
| 1(Gap(i,s,r)≥ 0) | 0.0050*** (0.0002) | 0.0124*** (0.0003) | 0.0171*** (0.0002) | 0.0151*** (0.0022) | 0.0992*** (0.0032) | 0.0666*** (0.0033) | 0.0025** (0.0012) | 0.0037** (0.0018) |
| First Stage - DepVar: 1(Gap(i,s,r)≥ 0 | | | | | | | | |
| 1(Others'Gap(i,s,r)≥ 0) | 0.1843*** (0.0003) | 0.2092*** (0.0008) | 0.1812*** (0.0003) | 0.1888*** (0.0012) | 0.2088*** (0.0020) | 0.1469*** (0.0012) | | |
| Wait Time | | | | | | | 0.0024*** (0.0000) | |
| Wait Time Others | | | | | | | | 0.0031*** (0.0000) |
| Observations | 69,625,780 | 69,625,780 | 69,625,780 | 5,096,109 | 5,096,109 | 5,096,109 | 4,969,924 | 5,131,022 |
| 1 st -Stage KP F-Stat | 333599.6 | 70567.1 | 411840.3 | 25200 | 10411.2 | 17412.6 | 8634.704 | 4850.91 |
| 1 st -Stage r2 | 0.6090 | 0.3002 | 0.3602 | 0.6104 | 0.3386 | 0.5722 | 0.2133 | 0.1923 |
| Shifts | 3,036,570 | 3,036,570 | 3,036,570 | 205,517 | 205,517 | 205,517 | 363,817 | 363,817 |
| Drivers | 12,451 | 12,451 | 12,451 | 12,445 | 12,445 | 12,445 | 12,103 | 12,103 |
| Fixed Effects | Shift-Driver | Driver | Driver-Day-Hour | Shift-Driver | Driver | Driver-Day-Hour | Driver | Driver |
| | | | | | | | Trip Number | Trip Number |
| | | | | | | | Month | Month |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the fixed effect level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day. Columns 1-6 present the IV based on Other Drivers' Income Gap, while column 7 shows results from an IV that uses waiting time at JFK as instrument.

While using other drivers to instrument for a given driver’s income gap addresses the concern that our estimates are solely driven by predetermined effort choices, there are some possible threats to the exclusion restriction. That is, the aggregate demand fluctuations underlying other drivers’ gaps may have a direct effect on a given driver’s cheating propensity, other than through his own income gap. For example, the composition of customers may be correlated with the level of demand (e.g. tourists may make up a larger proportion of customers when demand is low), and informational asymmetries are important determinants of fraud propensity (Balafoutas et al. 2012). To partially address this concern, we include ending hour and location fixed effects in all IV specifications. Thus, the exclusion restriction would only be violated if the passenger composition manifests in a way that is relatively idiosyncratic.

In Column 7 and 8 of Table 5 we show the result from a further attempt to isolate biases arising from the demand side, focusing our attention to the assignment of drivers to customers at John F. Kennedy International Airport (JFK) and La Guardia (LGA) airports. When a driver drops a customer off at the airport, he has a choice of whether to go back empty to the city or wait in a parking lot. Drivers who decide to stay must queue for a variable time period that depends on the intersection of demand (travellers) and supply (other drivers ahead in the queue). The Port Authority allocates the drivers to the several terminals to meet new customers. The queues of both passengers and drivers produce quasi-random variation in the match characteristics. We exploit this double source of variation in waiting time and passenger characteristics to identify the effect of longer waiting times on the propensity to commit fraud on subsequent rides. Specifically, we use the waiting time at the airport on ride r to instrument for whether the driver is above target on rides $r + 1$ to the last (the sample is limited to rides $r + 1$ to last).³⁰ By using variation in the income

³⁰We refer to the time between drop-off time of ride $r - 1$ and the pick-up time for ride r as waiting

gap on rides $r + 1$ to last derived from (possible) demand fluctuations at the airport on ride r , we isolate variation that is likely independent of predetermined effort choices of drivers and passenger composition (since passenger r characteristics are likely to be independent of passenger $r + 1$ characteristics).

The result reported in column 7 and 8 include several fixed effects: driver, trip number, and month (to account for seasonality). The regression include all rides after the airport – we are not able to find results when we restrict to only the first ride on which it is possible to cheat after the airport. Finally, the sample is restricted to drivers who were brought to the airport by a previous rides (did not go empty to the airport), and who decided to wait at the airport for the next sample. Aside from these selection issues, the results confirm the pattern observed in the other specifications. We report the full set of results in Appendix B.3.

2.5.4 Liquidity Constraints

Table 6 sheds some light on the possibility that liquidity constraints could operate at the daily level. We do so by presenting results on linear probability model regressions that include driver-shift fixed effects. Columns 1 presents a regression that includes realized earnings over the course of the shift in the regression. If drivers hold a stable expectation on the autocorrelation structure of cumulative income over a shift and income effects are negligible, then the inclusion of cumulative income would capture the distance between expected final earnings and daily targets for a liquidity constrained driver. This allows us to shed some light on whether the gap of current income from expectations operates through reference-

time if the location of both rides is in the airport census tract. While it is possible that some drivers took anticipated breaks during this period, there should be some random variation in how long drivers stayed within that location.

dependence rather than liquidity constraints story. The coefficient on the positive Income Gap indicator is unaffected by the inclusion of cumulative income. However, this could also happen if daily needs vary across shifts based on unobservables..

To further investigate the presence of liquidity constraints, we compute a measure of the income earned by each driver over the last four shifts in which he operated. We again divide the sample of shifts for each driver in quartiles. Columns 2-5 presents results for this split, from low income over the last four shifts (col. 2) to high income (col. 5). The heterogeneity reported shows that income in the very recent past, likely correlated to cash at hand, are correlated with a larger effect. In particular, we can observe two effects. The average propensity to cheat is larger when recent income have being lower, and the relative increase in this propensity to cheat when below target is also even larger. This result might be consistent with liquidity constraints, if recent income proxy for cash at hand. However, it is also consistent with a decrease in moral costs, if drivers feel that they are going through an unjust period of bad luck. Finally, if days of bad income make drivers more attentive to even small income fluctuations, these regressions are in agreement with our main hypothesis of reference-dependence and aversion to losses.

Table 2.6: LPM Evidence on Liquidity Constraints: Fraud Probability, by Ride

| | Income | Past Earnings | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | 1st quartile | 2nd quartile | 3rdquartile | 4th quartile |
| $\mathbf{1}(\text{Gap}(i,s,r) \geq 0)$ | 0.0006*** (0.0000) | 0.0009*** (0.0001) | 0.0007*** (0.0001) | 0.0005*** (0.0001) | 0.0004*** (0.0000) |
| Cumulative Income (Excludes Fraud)/100 | 0.0003** (0.0001) | | | | |
| Pr(Fraud) | 0.0084 | 0.0116 | 0.0089 | 0.0073 | 0.0062 |
| Observations | 73,033,879 | 15,055,505 | 17,251,713 | 18,756,657 | 20,869,916 |
| r ² | 0.3778 | 0.4411 | 0.4026 | 0.3551 | 0.3002 |
| Shifts | 3,074,836 | 761,116 | 754,890 | 757,891 | 751,554 |
| Drivers | 12,446 | 12,444 | 12,442 | 12,445 | 12,447 |
| Fixed Effects | Driver-Shift | Driver-Shift | Driver-Shift | Driver-Shift | Driver-Shift |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

2.6 Conclusion

Using field evidence from a widespread episode of fraud among New York City taxi drivers, this paper presents results consistent with the view that loss aversion motivates drivers to cheat. The effect is robust to various specifications and identification strategies. Our estimates are supportive of models of reference dependent preferences that take drivers expectation as reference points. We find that a driver is more likely to cheat more on shifts with lower income than expected and on those rides within a shift in which he is below our estimated expectations. We find that this relationship is more marked when the distance from expectation is larger (above \$25), possibly explained by drivers' inattentiveness to very

small fluctuations or by measurement error in the expectation term.

We provide results from instrumental variable regressions that use other drivers' earnings fluctuations as an instrument. This instrument allows us to more convincingly isolate the component of the deviation from expectation that is not due to effort adjustment. Using this methodology, we confirm a sizeable effect in the propensity to cheat when drivers' income realizations are unexpectedly low.

Two other complementary mechanisms could also play a role in explaining the field evidence. First, continuously binding liquidity constraints among New York City taxi drivers could lead to more fraud in response to deviations from the expected income accumulation path. While a sizeable effect size is observed also for small deviations, our heterogeneity analysis shows correlational evidence that very recent income affect the propensity to cheat in response to losses. This suggests that availability of cash at hand might play an important role in cheating decisions. However, this evidence is consistent with increased attention to income fluctuations. Moreover, we find that the effect of ride-level deviations from expectations is robust to the inclusion of cumulative income and driver-shift fixed effects in the regression, suggesting that controlling for time-invariant factors and variables related to the actual difficulty to meet daily needs does not affect our estimates.

A second complementary mechanism stems from the fact that drivers could perceive the deviation from the reference point as unjust from a metaphysical standpoint. We would then be identifying off a temporary reduction in moral costs, rather than loss aversion. Ultimately, the field evidence does not allow us to definitely rule out these two observationally equivalent mechanisms. Even in an experimental setting, it would be unclear how to isolate moral costs from loss aversion determinants and we do not know of any work that attempts to do so. One possibility would be to estimate a more complex structural model that relaxes the assumption of additive separability in moral costs. This is beyond the scope of our current

work, and would also require us to make stronger assumptions on functional forms in the realm of ethical decision making.

A caveat of our analysis is that we found more fraud violations than reported by the Taxi and Limousine Commission (TLC), although we used one of the procedures described by the Commission. It is possible that the TLC had access to other variables that were not provided in our data. The difference suggests that our results could be subject to measurement error and our estimates could suffer from attenuation bias. Even though measurement error could attenuate our estimates, we still find a sizeable effect of deviations from expectations on the propensity to commit an unethical action.

The observation that reference-dependence with respect expectations affects the perceived benefit of an offense enriches our understanding of ethical problems. Together with standard motivations, such as liquidity constraints and variations in moral costs, the evidence highlights the role that cognitive biases play in the assessment of the benefits of a crime. Our results suggest that psychological reactions might distort the cost benefit analysis of whether to commit a crime in systematic ways.

Chapter 3

Investment Incentives for Lagging Areas: Evaluating Firms' Responses to Incentive Subsidies using a Regression Discontinuity Approach

Giovanni Paci

3.1 Introduction

Regional development policies have gained momentum in the last few decades in both Europe and the United States. In the case of Europe, a designated agency, the European Structural Development Fund, has among its objectives the modernization and diversification of lagging regional industrial areas and the maintenance and creation of jobs, with an allocated budget of €278 billion for the period 2007-2013. Given the cost for taxpayers, several papers in the last decades have tried to fill the gap left by the absence of empirical evidence on regional policies. This paper presents evidence on one such program using a regression discontinuity design based on the assignment mechanism used by the Italian government.

The program examined here was launched by the Italian government in 1992. The Ministry of Industry financed project-related investment grants for about €16 billion over the period 1996-2003. The funded firms belonged to the manufacturing and extractive sectors and invested in territories designated as Objective 1, 2 and 5b by the European Community Structural Fund. The Ministry ranked the applications separately for each region based on several criteria such as the number of jobs involved. Within each region, the Ministry then funded firms from top to bottom based on their designated rank until the quota of funds allocated to each region was exhausted. This assignment generates a number of regional discontinuities where firms that were close in ranking but on either side of the cutoff obtained or failed to obtain the award. I pool these discontinuities to test whether firms that marginally won the award invested more in subsequent years than firms that marginally lost it.

Taking advantage of a dataset that was previously introduced by Bronzini and De Blasio (2006), I look at the performance of firms for a number of years after the first tranche of the

grant¹

I find that treated firms display an increase in investment over sales after receiving the grant. This effect is more marked among firms whose headquarters are in the South, for whom the grants represented a higher fraction of the total investment. However, averaging over the three years after the award, and in comparison with the three years before, there is no change in firms' performance on any other outcome. This could be due to a number of possibly complementary reasons. The projects financed could generate little profits or further investment, there could be time substitution, or the post-intervention period for the analysis could be too short. Moreover, the positive effects of externalities in production cannot be assessed using these data.

Using a difference in differences strategy, Branzini and De Blasio (2006) found a more marked increase in investment in the second year of the program followed by a decrease in later years, a pattern consistent with time substitution in investment strategy. Moreover, they discuss correlational evidence suggesting that subsidized firms show better performance on smaller markets, likely at the expense of non-subsidized firms. The regression discontinuity design I adopt in the analysis has three advantages over the difference-in-difference estimation in this setting. First, in the context of the assignment of investment grants, the firms' inability to precisely control the assignment variable near the cutoff generates a credible quasi-random variation even when the overall assignment mechanism is subject to unobservables. Second, firms were largely heterogeneous in regards to size, profit growth and cash flows before the auction year, suggesting that treated firms that received highest scores were likely to differ from all others also on unobservables, thereby confounding the differential impact on the treated firms discussed in previous work. Finally, treated and

¹The grant was disbursed in three tranches and received by the firms within three years from the award date.

untreated firms were likely to operate in different local markets and hence the assumption of common economic trends necessary for the difference-in-difference estimation is likely to be too strong.

The incentive program under scrutiny is part of a larger effort to promote the development of industrial clusters in the Italian depressed areas². There are several theoretical reasons that support policies designed to foster industrial agglomeration, whose benefits were first discussed in Marshall (1890). Clusters of firms facilitate the diffusion of information and technology, encourage the formation of a local labor force which possesses specific human capital for production, and reduce input costs by taking advantage of externalities that affect transportation costs. One of the main ideas behind regional policies is that, given the evidence that the location of firms is affected by agglomeration effects (Devereux, Griffith, and Simpson, 2007), the establishment of new firms can catalyze a cascade of local development. Moreover, policy makers hope that the increase in local production will support a stable local demand for services and amenities. Finally, programs promoting investment could also be beneficial if they sustain firms operating in areas with credit market imperfections.

However, some evidence supports a less encouraging depiction of the effects of these policies. A first issue, which has been shown using data collected from the Regional Selective Assistance in UK, is that these programs tend to support less efficient firms, which decreases aggregate productivity (Criscuolo, Matin, Overman, and Van Reenen, 2007). In the UK, even though there was a positive impact on firm employment and investment, the program supported less efficient firms and aggregate productivity growth had slowed down (a phenomenon referred to as cross-sectional substitution).

²A complementary government sponsored program for the same areas is the Patti Territoriali (Territorial Pacts, TPs). This program consists in an agreement between local governments, entrepreneurs and trade unions, subsequently endorsed by the central government, on a development plan for the area. The agreements include private and public investments provided by the central government.

Programs that subsidize reductions in tax margins to induce firms to relocate have likewise been ineffective. In the case of the European Structural Fund, Crozet, Mayer, and Mucchielli (2004) found little entry by firms. This could be related to offsetting general equilibrium effects such as increases in local land prices or to the fact that subsidies are small relative to agglomeration externalities in those areas and the high cost of relocation.

Martin, Mayners and Py (2012) report evidence consistent with these considerations in the context of the Zones Franches Urbaines (ZFU). These are areas where existing or new plants receive a broad array of tax subsidies ranging from social contribution payments to property taxes. The authors find that these policies are more effective when the cost of relocation is smaller, that these areas tend to attract smaller firms, and that the increase in local development is obtained by opportunistic relocation at the expense of non-subsidized areas.

Empirical support for an increase in local demand caused by the newly attracted labor force has been the subject of recent investigations (Chaury, 2014). The best evidence on this is from the U.S. Federal Empowerment Zone program, which focuses on spatially targeted tax incentives and grants. Busso, Gregory, and Kline (2010) report short-run increases in employment and wages but no influx of households to zone neighborhoods. They report evidence on rental prices that suggests workers consider zone neighborhoods poor substitutes for areas outside the subsidized zones.

Many commentators also note an inherently distortionary quality of these programs in that they encourage a culture of rent-seeking among entrepreneurs (Alesina et al., 2001) and can therefore increase corruption. In the case of the Italian Mezzogiorno, where judicial activity has clearly demonstrated the presence of criminal groups operating in the economy, this is certainly a primary consideration.

Finally, due to the complex nature of the market failures at work in each region, it is

unclear what kind of general rule would be effective to spur development. In Italy, the territories that constitute Objective 1 of the Regional Policies are those of the Italian Mezzogiorno regions. They include Sicily, Calabria, Basilicata, Apulia, Molise, Campania and Sardinia. These are regions historically characterized by profound disadvantages in education and welfare opportunities whose industrial production has consistently lagged behind the rest of the country in the last century (Malanima, 2005). The problem of developing the Italian Mezzogiorno has been a hotly debated topic among Italian scholars since the unification of the country and has focused on productivity measures, specific market failures, and historical reasons for the economic divergence of the South (Ciocca and Toniolo, 2004). Nevertheless, industrial policies are somewhat blind to the specific difficulties experienced by each of these territories.

The evaluation of local tax incentives is complicated by a variety of econometric challenges. For the study of the individual firm, a first problem is the absence of a relevant counterfactual that could allow the researcher to assess how much the firm would have invested in the absence of the subsidy. If programs tend to finance investment that would have happened even in the absence of the policy, taxpayers end up subsidizing firms. Similarly, if firms only make investments that they would have carried out in other years, policies are only inducing changes in the temporal patterns of investment. A second problem arises if the programs subsidize inefficient firms that would otherwise exit the market, and in this case the effects on the single firm are misleading on the aggregate. On the other hand, evaluations based on aggregate performance of these areas face the issue of selection bias. Areas that receive public support are inherently different from those areas that are excluded from it and are likely to be subject to unobserved time varying shocks.

In this context, this paper presents results for a localized treatment effect at the level of individual firms. By comparing firms who are close to the threshold, problems due to firm

entry and exit are arguably mitigated. This strategy allows me to ask simple questions such as whether investment, labor costs, and profit margins are increased by the grants. The answer is largely a negative one.

The paper proceeds as follows. Section 3.2 presents the institutional context and describes the data. Section 3.3 presents evidence from the regression discontinuity approach. Section 3.4 discusses heterogeneity in the treatment effects, and Section 3.5 concludes.

3.2 Institutional Context and Data

The data for this study were provided by the Bank of Italy and have been assembled by Branzino and De Blasio (2006), who linked administrative data obtained from the Ministry of Industry with financial statement information collected by the Bank. The data contain information on the firms that applied to subsidies under the Law 488 in rounds two (1997) and three (1998), the rank assigned by the Ministry to the firms, the amount of subsidy received, if any, and financial information for the years 1994-2001. I received two balanced panels that follow 1,007 and 1,329 firms for the second and third round of applications, respectively. Before describing the data, I present details on the institutional context.

3.2.1 Institutional Context

The Law 488/1992 was a major investment incentive program created in Italy for the promotion of industrialization in lagging areas.³ It took the form of project-related investment grants and disbursed subsidies for 27,846 projects in the period 1996-2003 for a total amount of approximately €16 billion.

³The Law was announced by the Italian government in 1992 (*Gazzetta Ufficiale della Repubblica Italiana* no. 299, 21 December 1992).

Eligible firms were those in manufacturing and extractive sectors⁴ that invested in areas designated as Objective 1 (GDP per capita less than 75% of European Union average), 2 (regions highly invested in declining industries), and 5b (peripheral rural regions). In addition to these three areas, the Ministry also decided to include areas that were excluded by the European Community Structural Fund but that had nonetheless been approved to receive grants pursuant to Article 92(3)c of the European Commission.⁵

A broad range of investment projects were eligible for grants, including extension, modernization, and restructuring of plants as well as those of relocation and rehabilitation for a new usage. The grants could not be combined with any other source of financing, and in order to maintain eligibility, any firm applying for the grant had to give up any other form of public assistance.

The applications were collected by the Ministry of Industry, which then ranked each region separately based on a number of predetermined criteria that were known beforehand by the firms. The maximum award amounts varied depending on the region and the size of the firms. For instance, for a large enterprise, the maximum award in areas classified as Objective 1 amounted to 50% of the investment. A firm was eligible for a single grant at each round but could apply to several rounds of the program, thereby receiving multiple grants.

For the first two rounds, the projects were ranked by the Ministry based on three criteria. A firm would get a positive score in proportion of the firm's own funds that were invested in the project and the number of jobs involved. It would get a negative score for the proportion of assistance that the firm sought in relation to the maximum allowed award rate. Starting in

⁴Selected producers in the tourism and transport sectors were included among the beneficiaries after 2001. They are excluded from the analysis.

⁵Article 92(3)c, which is now Article 87, defines eligible government aid compatible with the European common market.

the third round, the Ministry introduced a score that was based on regional priorities, which were determined by the central government, and a score based on environmental impact. The three (five) criteria had the same weight for the computation of the final score. Firms were ranked based on their scores within each region. Funds were distributed from the top to bottom ranked firms until the regional quota was exhausted. Each firm, if it received funding, would receive the rate that it requested in the application.

The grant was transferred from the Ministry to the awarded firms in three tranches of equal size. The first tranche was transferred within two months after the publication of the ranking (four months after the application deadline) and did not require the project to have been started. The second and third tranches were paid one and two years after the application deadline date, respectively. Contrary to the first tranche, the second and third tranches were received by the firms conditional on the project having advanced to at least two third of the final outcome. For projects whose durations were fewer than 24 months, there were only two tranches of equal size. In practice, for the second and third round of the program, the first tranche was paid with approximately a month of delay. The application deadline for the second round was February 1997 and the first tranche was received starting in July. The application deadline for the third round was also in February, but the first tranche was received by most firms in October.

3.2.2 Data Description

The data for this study was compiled by Bronzini and De Blasio (2006). It is constructed by linking the Law 488 dataset obtained by the Ministry of Industry with a dataset of firms' financial statements collected by the Bank of Italy, called Cerved. The Law 488 data contains information on all firms that submitted an application to the Ministry and the rank (but not the score) that each firm received. The Cerved data contains financial statements for a

large number of Italian corporations and covers 1993-2001. While the Cerved data makes it possible to study investment outcomes, it has the drawback of being representative only of large firms. Moreover, the linking procedure was complicated by a number of glitches and coding errors in both datasets.

The data I received are for the second and third rounds of the investment program. Starting from 3,358 (second round) and 3,751 (third round) applicant firms contained in the Ministry of Industry data, Bronzini and De Blasio discarded firms that received grants from the program more than once or that received grants in later rounds prior to 2001. They then linked the firms to the Cerved data using fiscal identifiers and Chamber of Commerce codes available in both datasets. The authors were able to link 1,196 (second round) and 1,498 (third round) firms. Finally, they selected firms with non-negative values of assets, capital stock, and sales and trimmed extreme observation on those variables, retaining only firms between the 5th and the 95th percentile. As a result of this cleaning procedure, they were left with two balanced panels, one for the second and one for the third round. The panel for the second round contains financial statements for 1,008 firms (1,007 firms in the file I received) for the years 1994-2001. The panel for the third round contains information on 1,329 firms for the years 1995-2001.

The variables used in each of the data sets are as follows: firm identifier, year, region, industrial sector, investment program round, rank obtained by the firm, whether the firm received a grant, size of the grant, size of the investment, sector classification (Ateco 1991), sales, value added, assets, liabilities and equity, interest costs, return on equity, total investments, leverage, cash flow, and cost of labor.

Although the dataset I received had been cleaned for a difference in differences estimation, it still contained a number of glitches regarding the main variables used in the analysis that follows. Given that regression discontinuity is sensitive to outliers close to the cutoff, I

excluded a number of firms. First, I excluded from the analysis 34 firms that were part of the treated group both in the second and in the third round of the program. Second, I excluded 3 firms located in the region Friuli in round 2 since all firms received grants in that occasion. Finally, I discarded some observations that displayed implausibly high values with respect to the main variables of analysis. In particular, aside from missing values there were 69 firms in the second round and 68 firms in the third round that recorded total investments in excess of two times the amount of sales for some years; 44 firms (26 in the second round and 18 in the third) that had profits in excess of sales; and 5 firms in each round that had labor costs in excess of twice the volume of sales. While these corrections are somewhat arbitrary, I believe they are necessary given the localized nature of the analysis. Overall, the exclusion of these outliers leaves 891 firms (161 excluded) for the second round and 1,220 firms (109 excluded) for the third round. Table 3.1 presents differences in program characteristics for the two rounds. Appendix A provides summary statistics by year and program round.

Table 3.1: Mean awards by program round

| | Means | | | |
|---------------------|------------------|------------------|----------|------------------|
| | Round 2 | Round 3 | diff | Total |
| Award Rate | .1922 (.2131) | .0869 (.1598) | .1053*** | .1313 (.1913) |
| Investment Project | 2,286,664 | 2,216,933 | 69,732 | 2,246,365 |
| Size | (4,754,078) | (11,115,720) | | (8,995,426) |
| Sales in auction | 17,972 | 22,146 | -4,173 | 20,384 |
| year (in thousands) | (73,226) | (156,157) | | (127,882) |
| Observations | 891 | 1220 | | 2111 |

Notes: The sample is limited to 891 firms for the second round and 1,220 firms for the third round.

3.3 Regression Discontinuity

3.3.1 Methodology

The applications submitted by the firms were ranked by the Ministry of Industry for each region separately and grants were assigned from top to bottom until the designated regional funds were exhausted. This assignment mechanism generated a collection of regional cutoffs which form the base of the regression discontinuity analysis. In the regression discontinuity setup, regressions take the following form:

$$Y_{irt} = \alpha \mathbb{I}(R_{ir} \geq 0) + \beta_1 p(R_{ir}) + \beta_2 \mathbb{I}(R_{ir} \geq 0) * q(R_{ir}) + X_{rt} \gamma + \epsilon_{irt} \quad (3.1)$$

Where R_{ir} is the rank relative to the cutoff for firm i in region r . The indicator $\mathbb{I}(R_{ir} \geq 0)$ takes value 1 if the firm is treated, and $p()$ and $q()$ are polynomials in the relative rank. The outcome Y_{irt} is a measure of performance of the firm over a certain time period t . It is important to notice that one of the assumptions of the regression discontinuity design (Thistlewaite and Campbell, 1960) is violated. The parameter α is a locally unbiased estimate of treatment effect under the assumption that the outcome of interest is a smooth function of the forcing variable, the relative rank in this case. If this assumption was satisfied, firms that scored just below the cutoff would be an adequate control group for those located just above, and the change in the outcome of interest could be attributed to the difference in treatment status. However, the rank is an ordinal measure, and hence the outcome, even if it was a smooth function of the underlying *score*⁶, cannot be a smooth function of the *rank*. While this is a crucial limitation of this study, pooling the discontinuity in the 17 regions for the second round and 18 regions for the third round is likely to attenuate this problem.

⁶All of the numerous attempts to obtain the score from administrative sources proved unsuccessful.

In the result section, I show that firms at the cutoff are similar with respect to sales, assets, cash flows and debt levels, suggesting that the distortion introduced by the use of the rank is limited.

3.3.2 Visual Evidence

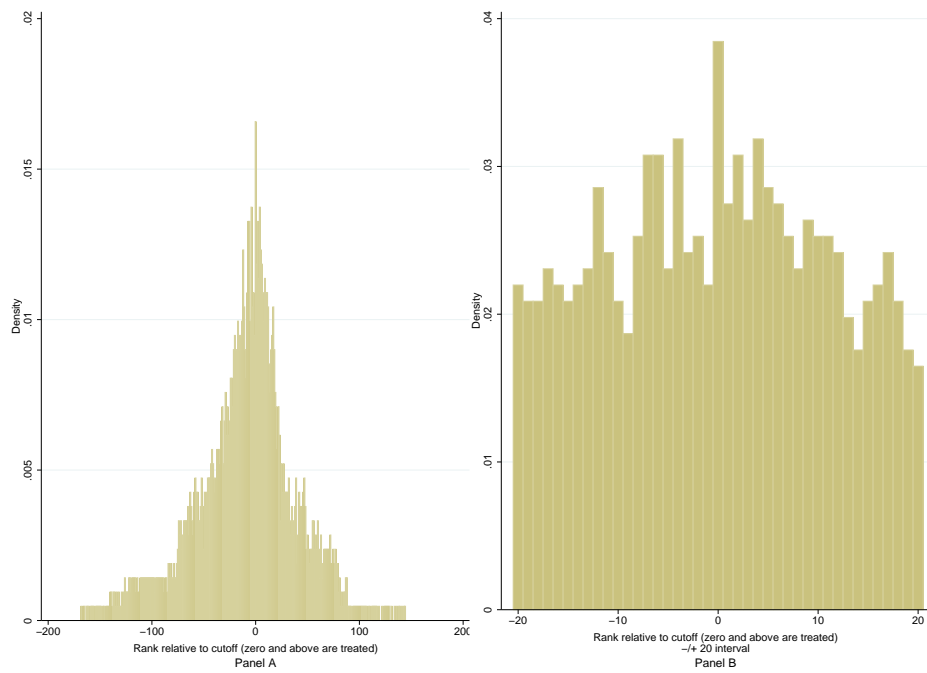
Figure 3.1 presents the histogram of the firms included in the sample by rank relative to the cutoff, for the full sample (Panel A) and for firms in a window of twenty rank positions relative to zero (Panel B). The award mechanism assigns one firm to each rank position for each of the two rounds and regions. Hence, if the number of participating firms was the same across regions and all the firms were in the sample, Panel A would show a uniform distribution. However, regions varied greatly in the number of firms participating and hence the histogram displays a larger mass around zero.

A McCrary (2008) type test for discontinuity in the density is not available for a discrete forcing variable, but Panel B of Figure 3.1 clearly demonstrates the bunching at zero (the lowest ranked firm to receive a grant). This evidence does not necessarily correspond to manipulation of the forcing variable since firms could not get a higher rank in any way and the Ministry published, by region, the full list of firms awarded and not. It could be due instead to missing or erroneous observations: firms that scored below the cutoff are less likely to be included in the sample (either as a consequence of the linking procedure with Cerved data or for the presence of outliers). This problem would hence be analogous in its effect to differential attrition.

A second possibility is that firms had some influence on the amount of funds available by regions, perhaps by lobbying. For instance, larger firms at the margin of winning the award could have been more effective in influencing the Ministry to increase the regional funds. Since the matching procedure with the Cerved data tended to favor the inclusion of

larger firms, this type of unobserved behavior by firms could generate a histogram like the one reported in Panel B. This would invalidate the RD approach since treated firms at the margin would be different based on unobservables (e.g. lobbying activity). However, to the best of my knowledge, there is no evidence that such pressure by firms on allocated funds has occurred in practice.

Figure 3.1: Histogram of relative rank

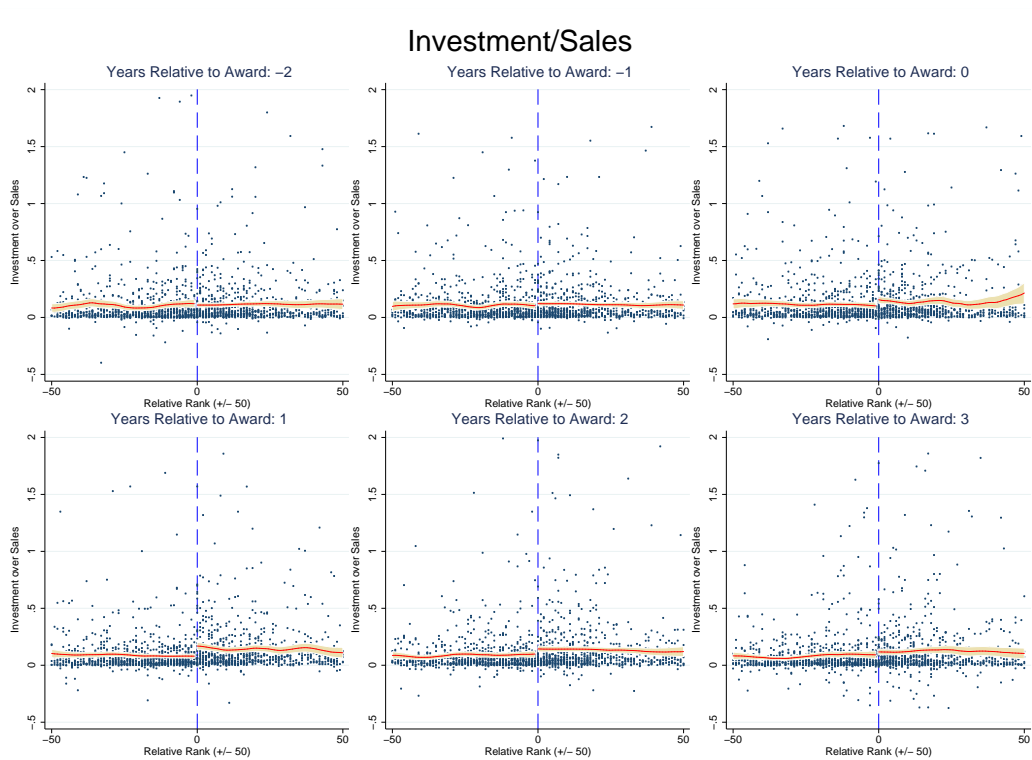


Notes: Panel A displays the histogram for the full sample. Panel B is restricted to firms whose position is within 20 places from the cutoff.

With this caveat in mind, the following figures report visual evidence on the discontinuity for the main outcome variable of interest. The graph presents a local polynomial regression that is allowed to vary at either side of the cutoff for investment as a fraction of sales. The cutoff is normalized so that zero and positive bins correspond to firms awarded the grant. The normalization of investment relative to sales is necessary as firms differed considerably

in size. The panels display the patterns of the outcome variable for the three years before and after the intervention. While the graph shows an increase in investment at time of the intervention, the increase is mostly reabsorbed by the last period. I run similar graphs for the profit margin and labor cost as fraction of sales in Appendix B. For both those variables, the graphical evidence does not suggest large changes in firm performance.

Figure 3.2: Local polynomial regressions for investment/sales for three years before and after the award



Notes: Local polynomial regressions by rank position relative to cutoff. Standard errors in shaded areas. The dots represent means of the y-axis variable for each position in the relative rank.

3.3.3 Regression Analysis

In this section, I report estimates from regressions that pool the 2,111 firms in the second and third round of the intervention. I split the sample in two blocks, one for the two years before

the award and the year in which the award is obtained and one for the three years after the award. With this cut of the data, each firm is observed over an equal number of years before and after participation in the program. Since at least two thirds of the investment had to be completed in order to receive the second tranche of grant, I refer to the second block as the post-intervention period.

Table 3.2 presents results for sales, investment over sales, profit margin, and labor costs over sales. Odd columns report results for the period before the intervention and even columns report results for the post-intervention period, similarly to Figure 3.2. Each regression controls for a third order polynomial in the rank that is allowed to vary at either side of the cutoff. The standard errors are clustered at each bin, in order to allow for random specification errors due to the discrete nature of the forcing variable. The results report an increase in investment for treated firms in the years following the intervention. This is to be expected if the funds received from the Ministry are indeed spent (or counted as) investment. For the other variables, the data do not suggest large effects. In particular, treated firms do not seem to increase their profit margin and sales are not affected.

Table 3.2: Regression discontinuity estimates for the main outcomes of interest, by pre- and post-award year

| | Sales (in thousands) | Sales (in thousands) | Inv./Sales | Inv./Sales | Profit/Sales | Profit/Sales | Lab. Cost/Sales | Lab. Cost/Sales |
|------------------------|-------------------------|-------------------------|------------|------------|--------------|--------------|-----------------|-----------------|
| | Before | After | Before | After | Before | After | Before | After |
| Treated | 13965.307 | 16151.785 | 0.019* | 0.058*** | -0.010 | -0.007 | -0.030* | -0.027* |
| (=1 if firm is funded) | (9762.809) | (10551.093) | (0.010) | (0.015) | (0.019) | (0.018) | (0.017) | (0.016) |
| Constant | 14143.289*** | 17949.517*** | 0.110*** | 0.089*** | 0.326*** | 0.309*** | 0.217*** | 0.215*** |
| | (3189.859) | (3636.968) | (0.006) | (0.007) | (0.015) | (0.014) | (0.013) | (0.012) |
| N | 6,333 | 6,333 | 6,333 | 6,333 | 6,333 | 6,333 | 6,333 | 6,333 |
| r ² | 0.008 | 0.008 | 0.003 | 0.015 | 0.005 | 0.002 | 0.005 | 0.003 |
| MeanDepVar | 19062.9032 | 23153.1420 | 0.1136 | 0.1065 | 0.3175 | 0.3036 | 0.1928 | 0.1955 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the rank relative to cutoff level in parenthesis. All specifications include a third order polynomial whose coefficients are allowed to vary at each side of the cutoff. The sample includes 2,111 firm for the two rounds observed in the three years before and after the year of application.

Instead of comparing firms that just won or lost the award by a few rank positions before and after the intervention, it is possible to run a regression discontinuity for the change in outcome variables before and after the intervention. The identifying assumption for this regression discontinuity is that firms on either side of the cutoff have similar trends, so that deviations can be attributed to the change in treatment status. Table 3.3 reports the main results from this regression approach. Even though investment is still marginally significant as a fraction of sales, overall investment over capital decreases equally among treated and non treated firms. While it is possible that treated firms anticipated investment they would have done in the second or third year, it is also possible that the grant substituted for investment that would have happened anyway. Moreover, because the general patterns are negative, the evidence is suggestive that the funded projects did not generate novel profit opportunities. Columns 6-8 report results on three financial variables. Treated firms seem to experience an increase in cash flows. Even though the evidence does not draw a clear picture, treated firms seem to be financially advantaged in the post-intervention period. This financial distortion,

however, does not translate into a substantive change in business, at least for the period of the intervention.

One limitation of the analysis is related to length of the sample. The investments funded by the Ministry could have a longer term impact that could not be picked up by the short post-intervention period. However, it seems plausible that the incentive was mostly used to finance investment opportunities that would have been taken anyway or that were largely unprofitable.

Table 3.3: Regression discontinuity estimates, post- pre-award differences

| | Δ Sales | Δ Inv./Sales | Δ Inv./Capital | Δ Profit/Sales | Δ Lab. Cost/Sales | Δ Assets | Δ Debt | Δ Cash Flow |
|------------------------|----------------|---------------------|-----------------------|-----------------------|--------------------------|-----------------|---------------|--------------------|
| Treated | 1985.826 | 0.035** | -0.394 | 0.002 | 0.004 | 5572.961 | 1556.117 | 2046.067* |
| (=1 if firm is funded) | (3457.018) | (0.016) | (0.545) | (0.009) | (0.006) | (5061.498) | (2353.587) | (1231.524) |
| Constant | 3972.840*** | -0.018** | -0.720** | -0.016*** | -0.002 | 4768.385*** | 3043.451** | 99.069 |
| | (1179.164) | (0.009) | (0.318) | (0.005) | (0.004) | (1793.145) | (1534.347) | (221.469) |
| N | 2,111 | 2,111 | 2,111 | 2,111 | 2,111 | 2,111 | 2,111 | 2,111 |
| r ² | 0.008 | 0.011 | 0.004 | 0.006 | 0.004 | 0.010 | 0.010 | 0.005 |
| MeanDepVar | 4090.2387 | -0.0071 | -0.4521 | -0.0140 | 0.0027 | 5017.3777 | 2939.8767 | 822.3944 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the rank relative to cutoff level in parenthesis. All specifications include a third order polynomial whose coefficient are allowed to vary at each side of the cutoff. The sample includes 2,111 firms for the two rounds observed in the three years before and after the year of application. The variables are average differences between the three year post-intervention period and the three year pre-intervention period.

3.4 Heterogeneity

Italy is a highly heterogeneous country. Its regions have different economic environments and firms differ in size and business culture.⁷ While most of the territories that qualify for structural funds are located in the South, the firms that participate in the Ministry of Industry investment program under study have headquarters all over the nation. It is

⁷See, for instance, Trigilia and Burroni (2009).

instructive to know whether the region to which the firm belongs has any effect on the effectiveness of the program. Because the assignment rule generates discontinuity at regional level, it could be that difference in firms size mask the effect of the program.

Table 3.4 reports results for firms belonging to the Central and Northern regions. For these firms there is a surprising dearth of effects, even when looking at investment over sales. This effect is related to firm size, as the investment from the grant was a relatively small share of the overall investment for larger firms. The effect of the program on investment is more marked in the Southern regions. A number of caveats apply to these results. In particular, it could be that for firms based in the Northern regions it takes longer to activate investment in lagging areas, which are often in the South. Ultimately, the shortness of the post-intervention period prevents me from making a definitive claim of ineffectiveness.

Table 3.4: Regression discontinuity estimates, post- pre-award differences by geography

| Panel A: Northern and Central Regions | | | | | | | | |
|---------------------------------------|----------------|---------------------|-----------------------|-----------------------|--------------------------|-----------------|---------------|--------------------|
| | Δ Sales | Δ Inv./Sales | Δ Inv./Capital | Δ Profit/Sales | Δ Lab. Cost/Sales | Δ Assets | Δ Debt | Δ Cash Flow |
| Treated | -1993.811 | 0.005 | -0.350 | -0.007 | -0.007 | -742.738 | -748.578 | 121.670 |
| (=1 if firm is funded) | (4131.963) | (0.020) | (1.143) | (0.012) | (0.007) | (3642.509) | (2825.528) | (959.106) |
| Constant | 4249.927** | -0.029** | -1.054* | -0.010 | -0.005 | 6807.598** | 4967.254** | 143.324 |
| | (1637.786) | (0.014) | (0.561) | (0.007) | (0.005) | (2744.527) | (2146.440) | (339.380) |
| N | 1,053 | 1,053 | 1,053 | 1,053 | 1,053 | 1,053 | 1,053 | 1,053 |
| r ² | 0.019 | 0.003 | 0.005 | 0.005 | 0.004 | 0.016 | 0.012 | 0.004 |
| MeanDepVar | 4056.9367 | -0.0224 | -0.6557 | -0.0168 | -0.0024 | 4103.4516 | 2297.2240 | 420.7513 |
| Panel B: Southern Regions | | | | | | | | |
| | Δ Sales | Δ Inv./Sales | Δ Inv./Capital | Δ Profit/Sales | Δ Lab. Cost/Sales | Δ Assets | Δ Debt | Δ Cash Flow |
| Treated | 3610.815 | 0.085*** | -0.324 | 0.008 | 0.012 | 12189.254 | 2296.023 | 3266.927 |
| (=1 if firm is funded) | (8489.106) | (0.030) | (0.434) | (0.019) | (0.014) | (11926.536) | (4470.700) | (3041.033) |
| Constant | 4215.179** | -0.006 | -0.237 | -0.024* | 0.001 | 3703.887*** | 2183.462** | 227.132 |
| | (1824.228) | (0.020) | (0.161) | (0.012) | (0.011) | (1234.287) | (1007.036) | (247.952) |
| N | 1,058 | 1,058 | 1,058 | 1,058 | 1,058 | 1,058 | 1,058 | 1,058 |
| r ² | 0.006 | 0.032 | 0.004 | 0.010 | 0.010 | 0.010 | 0.011 | 0.007 |
| MeanDepVar | 4123.3834 | 0.0082 | -0.2494 | -0.0111 | 0.0077 | 5926.9847 | 3579.4923 | 1222.1394 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the rank relative to cutoff level in parenthesis. All specifications include a third order polynomial whose coefficient are allowed to vary at each side of the cutoff. The sample includes firms whose headquarters are in the Central or Northern regions of Italy. Variables represent average differences for the two rounds observed in the three years before and after the year of application.

A second limitation of the heterogeneity analysis stems from the fact that the data are at the firm rather than plant level. If firms in the Northern Italy have production plants across the country, it is possible that the absence of increase in investment masks an underlying regional reallocation. The program could induce a larger investment in lagging areas at the expense of other areas. The absence of data on this type of geographical redistribution remains an important limitation of the firm level analysis.

3.5 Conclusion

Investment incentives programs for in lagging regions are an important policy tool in Europe and the United States. Evaluating such programs has a number of difficulties ranging from the availability of suitable data to econometric complications.

In this paper I study a program funded by the Italian Government in 1992. The Ministry of Industry financed project-related investment grants for about €16 billion from 1996-2003 for firms investing in territories designated as Objective 1, 2 and 5b by the European Community Structural Fund. The assignment mechanism used by the Ministry allows me to take advantage of a pool of regression discontinuities at the regional level. Within each region, the Ministry ranked firms' projects and then funded them from top to bottom until the quota of funds allocated to the region was exhausted. I pool these discontinuities to test whether firms that marginally won the award invested more in subsequent years than firms that lost it.

My results are complementary to those obtained by Bronzini and De Blasio (2006), who looked at the effect of the program using a difference in differences on the same data. In the context of the assignment of investment grants, the firms inability to precisely control the assignment variable near the cutoff generates a credible quasi-random variation even when the overall assignment mechanism is subject to unobservables. The RD approach reveals a starker absence of responses by firms than the difference-in-difference, suggesting that the assumption of common trends made in previous work might not be appropriate in this context.

Treated firms display an increase in investment over sales after receiving the grant. This effect is more marked among firms whose headquarters are in the South and for whom the grants formed a higher fraction of the total investment. However, there is no change in firms'

performance on any other outcome. The projects financed could not be very profitable or could not generate further investment. Moreover, there could be a pattern of time substitution if firms anticipate investment that they would have sustained anyway in future years. It is also possible that the post-intervention period for the analysis used in the analysis is too short. I look at changes in the three years after the award is received, and it is possible that the chosen investment became productive over a longer time span.

The evaluation of an investment incentive program at level of the individual firm does not do justice to the nature of the program. The program may be designed to foster industrial agglomeration that facilitates the diffusion of information and technology, creates a local labor force, and develops local input cost externalities, and studying the effects on individual firms would not fully identify effects. In order to assess the benefits of such programs, the feedback on the individual firms of these externalities might take longer to realize, and the effects on other untreated firms are the relevant ones. The data at hand does not allow me to tackle this question, but it should be the object of future enquiry in the Italian context given the long history of distortionary effects of public funding on the local economies.

Bibliography

- [1] Alberto Abadie and Sebastien Gay. The impact of presumed consent legislation on cadaveric organ donation: a cross-country study. *Journal of health economics*, 25(4):599–620, July 2006.
- [2] Alberto Alesina, Edward Glaeser, and Bruce Sacerdote. Why Doesn’t the United State Have a European-Style Welfare State? *Brookings Papers on Economic Activity*, 2:187–254, 2001.
- [3] Ofer H. Azar. Do people tip because of psychological or strategic motivations? An empirical analysis of restaurant tipping. *Applied Economics*, 42(23):3039–3044, September 2010.
- [4] Ofer H. Azar. Business strategy and the social norm of tipping. *Journal of Economic Psychology*, 32(3):515–525, June 2011.
- [5] Loukas Balafoutas, Adrian Beck, Rudolf Kerschbamer, and Matthias Sutter. What drives taxi drivers? A field experiment on fraud in a market for credence goods. *Review of Economic Studies*, 2013.
- [6] Abhijit V Banerjee and Sendhil Mullainathan. Limited Attention and Income Distribution. *The American Economic Review*, 98(2):489–493, April 2008.

- [7] Nicholas C. Barberis. Thirty Years of Prospect Theory in Economics: a Review and Assessment. *NBER working paper 18621*, 2012.
- [8] Gary S. Becker. Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2):169–217, 1968.
- [9] John Beshears. Active Choice and Health Care Costs : Evidence from Prescription Drug Home Delivery. *Working Paper*, pages 1–44, 2012.
- [10] John Beshears, James J Choi, and Brigitte C Madrian. The Limitations of Defaults. *Working Paper*, 2010.
- [11] Erin Todd Bronchetti, Thomas S Dee, David B Huffman, and Ellen Magenheimer. When a Nudge isn’t Enough: Defaults and Saving among Low-Income Tax Filers. *NBER working paper 16887*, 2011.
- [12] Raffaello Bronzini and Guido de Blasio. Evaluating the impact of investment incentives: The case of Italy’s Law 488/1992. *Journal of Urban Economics*, 60(2):327–349, September 2006.
- [13] Matias Busso, Jesse Gregory, and Patrick Kline. Assessing the Incidence and Efficiency of a Prominent Place Based Policy. *American Economic Review*, 103(2):897–947, April 2013.
- [14] Colin Camerer, Linda Babcock, George Loewenstein, and Richard Thaler. Labor Supply of New York City Cabdrivers: One Day at a Time. *The Quarterly Journal of Economics*, 112(2):407–441, 1997.

- [15] D. Card and G. B. Dahl. Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior. *The Quarterly Journal of Economics*, 126(1):103–143, March 2011.
- [16] Gabriel D Carroll, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick. Optimal Defaults and Active Decisions. *The Quarterly Journal of Economics*, 124(4):1639–1674, November 2009.
- [17] Marco Castillo, Ragan Petrie, Maximo Torero, and Lise Vesterlund. Gender differences in bargaining outcomes: A field experiment on discrimination. *Journal of Public Economics*, 99:35–48, March 2013.
- [18] M Keith Chen, Venkat Lakshminarayanan, and Laurie R. Santos. How Basic Are Behavioral Biases? Evidence from Capuchin Monkey Trading Behavior. *Journal of Political Economy*, 114(3), 2006.
- [19] Pierluigi Ciocca and Gianni Toniolo. *Storia economica d’Italia*. Laterza, 2004.
- [20] Vincent P. Crawford and Juanjuan Meng. New York City Cab Drivers’ Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. *The American Economic Review*, 101(August):1912–1932, 2011.
- [21] Chiara Criscuolo, Ralf Martin, and Henry Overman. The effect of industrial policy on corporate performance : Evidence from panel data. *Working Paper*, pages 1–49, 2007.
- [22] Henrik Cronqvist and Richard H. Thaler. Design Choices in Privatized Social-Security Systems : Learning from the Swedish Experience. *The American Economic Review*, 94(2), 2004.

- [23] Michael R. Cunningham. Weather, mood, and helping behavior: Quasi experiments with the sunshine samaritan. *Journal of Personality and Social Psychology*, 37(11):1947–1956, 1979.
- [24] Vittorio Daniele and Paolo Malanima. Il prodotto delle regioni e il divario Nord-Sud in Italia (1861-2004). *Rivista di Politica Economica*, 2004.
- [25] Sibilla Di Guida, Davide Marchiori, and Ido Erev. Decisions among defaults and the effect of the option to do nothing. *Economics Letters*, 117(3):790–793, December 2012.
- [26] Kirk Bennett Doran. Wages, Daily Income Goals and Daily Labor Supply. *Working Paper*, pages 1–25, 2010.
- [27] Julie S Downs, George Loewenstein, and Jessica Wisdom. Strategies for Promoting Healthier Food Choices. *The American Economic Review*, 99(2):159–164, April 2009.
- [28] Gilles Duranton. California Dreamin’: The Feeble Case for Cluster Policies. *Working Paper*, 3:3–45, 2011.
- [29] Eurostat. Europe in Figures - Eurostat Yearbook 2009. 2009.
- [30] Henry S. Farber. Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers. *Journal of Political Economy*, 113(1):46–82, February 2005.
- [31] Henry S. Farber. Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. *The American Economic Review*, 98(3):1069–1082, 2008.
- [32] Ernst Fehr and Lorenz Goette. Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. *The American Economic Review*, 97(1), 1999.

- [33] Stephen M Fleming, Charlotte L Thomas, and Raymond J Dolan. Overcoming status quo bias in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):6005–9, March 2010.
- [34] David Genesove and Christopher Mayer. Loss Aversion and Seller Behavior: Evidence from the Housing Market. *The Quarterly Journal of Economics*, 116(4):1233–1260, 2001.
- [35] Jacob Goldin and Tatiana Homonoff. Smoke Gets in Your Eyes : Cigarette Tax Salience and Regressivity. *American Economic Journal: Economic Policy*, 5(1), 2013.
- [36] Kareem Haggag, Brian Mcmanus, and Giovanni Paci. Learning By Driving - Productivity Improvements by New York City Taxi Drivers. *Working Paper*, 2014.
- [37] Kareem Haggag and Giovanni Paci. Default Tips. *American Economic Journal: Applied Economics*, (Forthcoming).
- [38] Kirabo C. Jackson and Henry S. Schneider. Do Social Connections Reduce Moral Hazard? Evidence from the New York City Taxi Industry. *American Economic Journal: Applied Economics*, 3(July):244–267, 2011.
- [39] Eric J Johnson, Steven Bellman, and Gerald L Lohse. Defaults , Framing and Privacy : Why Opting In-Opting Out. *Marketing Letters*, 13(1):5–15, 2002.
- [40] Eric J Johnson and Daniel Goldstein. Do Defaults Save Lives? *Science*, 302:4–5, 2003.
- [41] Eric J Johnson, John Hershey, Jacqueline Meszaros, and Howard Kunreuther. Framing, Probability Distorsions, and Insurance Decisions. *Journal of Risk and Uncertainty*, 7:35–51, 1993.
- [42] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292, 1979.

- [43] Dean Karlan, John A List, and Eldar Shafir. Small Matches and Charitable Giving: Evidence from a Natural Field Experiment. *Journal of Public Economics*, 95(5-6):344–350, 2011.
- [44] Punam Anand Keller, Bari Harlam, George Loewenstein, and Kevin G. Volpp. Enhanced active choice: A new method to motivate behavior change. *Journal of Consumer Psychology*, 21(4):376–383, October 2011.
- [45] Daniel E. Keniston. Bargaining and Welfare: A Dynamic Structural Analysis. *Working Paper*, 2011.
- [46] Mary C Kern and Dolly Chugh. Bounded ethicality: the Perils of Loss Framing. *Psychological Science*, 20(3):378–84, March 2009.
- [47] Botond Koszegi and Matthew Rabin. A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, CXXI(4):1133–1165, 2006.
- [48] Botond Koszegi and Matthew Rabin. Reference-Dependent Risk Attitudes. *The American Economic Review*, 97(4):1047–1073, 2007.
- [49] Botond Koszegi and Matthew Rabin. Reference-Dependent Consumption Plans. *The American Economic Review*, 99(3):909–936, 2009.
- [50] David S. Lee and David Card. Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2):655–674, February 2008.
- [51] George Loewenstein and Scott Rick. Commentaries and Rejoinder to “The Dishonesty of Honest People”. *Journal of Marketing Research*, XLV(December), 2008.
- [52] Annamaria Lusardi. Saving and the Effectiveness of Financial Education. *Working Paper*, 2004.

- [53] Annamaria Lusardi and Olivia S. Mitchell. Financial literacy and retirement preparedness: Evidence and implications for financial education programs. *CFS Working Paper*, 2007.
- [54] Brigitte C Madrian and Dennis F. Shea. The Power of Suggestion: Inertia in 401(k) Participation and Saving Behavior. *The Quarterly Journal of Economics*, CXVI(November), 2001.
- [55] Paolo Malanima. Urbanisation and the Italian economy during the last millennium. *European review of economic history*, 9:97–122, 2005.
- [56] Ulrike Malmendier and Stefano Della Vigna. Paying Not to Go to the Gym. *The American Economic Review*, 96(3), 2006.
- [57] Philippe Martin, Thierry Mayer, and Florian Mayneris. Public Support to Clusters: A Firm Level Study of French "Local Productive Systems". *CEPR - Discussion Paper*, (7102), 2008.
- [58] Alexandre Mas. Pay, Reference Points, and Police Performance. *The Quarterly Journal of Economics*, CXXI(3), 2006.
- [59] Thierry Mayer, Florian Mayneris, and Loriane Py. The Impact of Urban Enterprise zones on Establishment's Location Decisions: Evidence from French ZFUs. *CEPR - Discussion Paper*, (9074), 2012.
- [60] Nina Mazar, On Amir, and Dan Ariely. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6):633–644, December 2008.

- [61] Justin McCrary. Manipulation of the Running Variable in the Regression Discontinuity Design: a Density Test. *Journal of Econometrics*, 142(2):698–714, 2008.
- [62] Terrance Odean. Are Investors Reluctant to Realize Their Losses? *The Journal of Finance*, LIII(5):1775–1798, 1998.
- [63] Devin G. Pope and Maurice E. Schweitzer. Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes. *The American Economic Review*, 101(1):129–157, 2011.
- [64] Alex Rees-jones. Loss Aversion Motivates Tax Sheltering: Evidence From U.S. Tax Returns. *Working Paper*, 2013.
- [65] Schaller Consulting. The New York City Taxicab Fact Book. (March), 2006.
- [66] Henry Schneider. Moral Hazard in Leasing Contracts : Evidence from the New York City Taxi Industry. *Journal of Law and Economics*, 53(4):783–805, 2010.
- [67] Maurice E. Schweitzer, Lisa Ordonez, and Douma Bambi. Goal Setting as a Motivator of Unethical Behavior. *Academy of Management Journal*, 47(3):422–432, 2004.
- [68] Eesha Sharma, Nina Mazar, Adam L. Alter, and Dan Ariely. Financial Deprivation Selectively Shifts Moral Standards and Compromises Moral Decisions. *Working Paper*, pages 1–57, 2013.
- [69] Kelly Shue and Erzo F. P Luttmer. Who Misvotes? The Effect of Differential Cognition Costs on Election Outcomes. *American Economic Journal: Economic Policy*, 1(1):229–257, January 2009.
- [70] Itamar Simonson. Choice of Based and on Reasons : The Case Effects Attraction Compromise. *Journal of Consumer Research*, 16(2):158–174, 2013.

- [71] Richard Thaler, Amos Tversky, Daniel Kahneman, and Alan Schwartz. The Effect of Myopia and Loss Aversion on Risk Taking: an Experimental Test. *The Quarterly Journal of Economics*, 112(2):647–661, 1997.
- [72] Donald L. Thistlethwaite and Donald T. Campbell. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51:309–317, 1960.
- [73] Carlo Trigilia and Luigi Burroni. Italy: rise, decline and restructuring of a regionalized capitalism. *Economy and Society*, 38(4):630–653, November 2009.
- [74] Cyrus R. Vance. District Attorney Vance announces charges against 59 NYC taxicab drivers engaged in widespread fraud of customers. *New York County Office District Attorney’s Office*, 2010.

Appendices

Appendix A

Appendix for Chapter 1

A.1 Passenger Information Monitor Screen Examples

A.2 Supplement to Section 2 (Regression Discontinuity)

We re-estimate regression specification (1), changing the outcome variable to ones that should not be significantly affected by the default suggestions. Table B.1 shows that the treatment effects are small and in conflicting directions for ride distance and duration, and the effects on passenger count, hour of the day, and day of the week are insignificant.

Figure A.1: Passenger Information Monitors (PIM): Vendor (Top) & Competitor (Bottom)



Notes: The top and bottom rows present examples of screens from the Vendor and Competitor, respectively. The sequence of payment screens follows from left to right. The top row screens (Vendor) were taken from an NYC taxi driver training video uploaded to YouTube on October 22, 2010 (http://www.youtube.com/watch?v=VnciL_4g8CE). The bottom row screens (Competitor) were photographed by the authors in October 2012. The Competitor offered buttons of 20%, 25%, and 30% during the photographed period (2012); however, during the period of study (2009), the Competitor offered 15%, 20%, and 25% buttons.

Figure A.2: Vendor PIM Examples (above and below \$15)



Notes: The Vendor offered defaults of \$2/\$3/\$4 for amounts less than \$15 (left image) and defaults of 20%/25%/30% (computed on the fare) for fares greater than \$15 (right image). Images taken from an NYC taxi driver training video uploaded to YouTube on October 22, 2010 (<http://www.youtube.com/watch?v=VnciL4g8CE>)

Table A.1: Regression discontinuity for trip distance, ride duration, hour of pick-up, day of the week, and passenger count

| | (1) | (2) | (3) | (4) | (5) |
|------------------------------|----------------------|---------------------|-------------------|------------------|------------------|
| | Distance | Duration | Hour of Pick-Up | Day of the Week | Passenger Count |
| $\mathbb{1}(Fare_r \geq 15)$ | -0.015*** (0.005) | 0.147*** (0.024) | -0.011 (0.012) | 0.009 (0.009) | 0.003 (0.002) |
| N | 6,218,196 | 6,218,196 | 6,218,196 | 6,218,196 | 6,218,196 |
| r ² | 0.884 | 0.780 | 0.378 | 0.059 | 0.901 |
| DepVarMean | 4.380 | 20.894 | 11.612 | 3.231 | 2.000 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at each fare value (\$0.40 intervals), in parentheses. $\mathbb{1}(Fare_r \geq 15)$ is an indicator function that the fare is greater than or equal to \$15. Ride duration is constructed as the difference between the drop-off time and the pick-up time (in minutes). Trip distance is recorded in miles. DepVarMean is the mean of the dependent variable on rides with fares of \$14.90. All specifications include fixed effects for driver, pick-up day of the week, pick-up hour, pick-up location borough, and drop-off location borough. The sample is limited to fares between \$5 and \$25 on Vendor-equipped cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday).

Table B.2 shows heterogeneous treatment effects by the number of passengers in the taxi. We find similar effect sizes across the range of passenger count sub-samples.

Table A.2: Heterogeneity by Number of Passengers: Regression Discontinuity estimates of Default Effect on Tip Amount

| | (1) | (2) | (3) | (4) |
|------------------------------|---------------------|---------------------|---------------------|---------------------|
| | One | Two | Three | Four |
| $\mathbb{1}(Fare_r \geq 15)$ | 0.299*** (0.006) | 0.291*** (0.012) | 0.267*** (0.024) | 0.320*** (0.048) |
| N | 3,911,666 | 851,301 | 241,251 | 85,977 |
| r ² | 0.208 | 0.225 | 0.206 | 0.256 |
| DepVarMean | 2.206 | 2.251 | 2.284 | 2.249 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

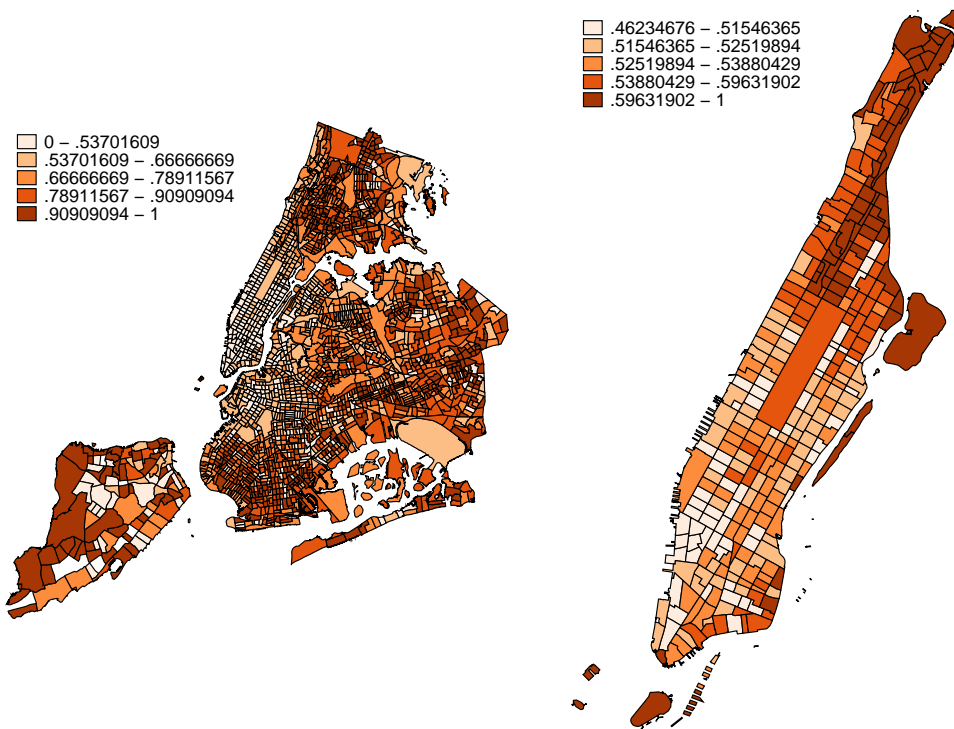
Notes: Robust standard errors clustered at each fare value (\$0.40 intervals), in parentheses. Columns (1) - (4) present the effect on tip amount by sub-samples of the number of passengers in the taxi. $\mathbb{1}(Fare_r \geq 15)$ is an indicator function that the fare is greater than or equal to \$15. DepVarMean is the mean of the dependent variable on rides with fares of \$14.90. All specifications include fixed effects for driver, pick-up day of the week, pick-up hour, pick-up location borough, and drop-off location borough. The sample is limited to fares between \$5 and \$25 on Vendor-equipped cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday)

A.3 Supplement to Section 3 (Comparing Across Vendors)

Figure C.1 shows the proportion of pick-ups that were on Vendor-equipped (relative to Competitor-equipped) taxis, within each census tract in Manhattan and New York City more broadly.

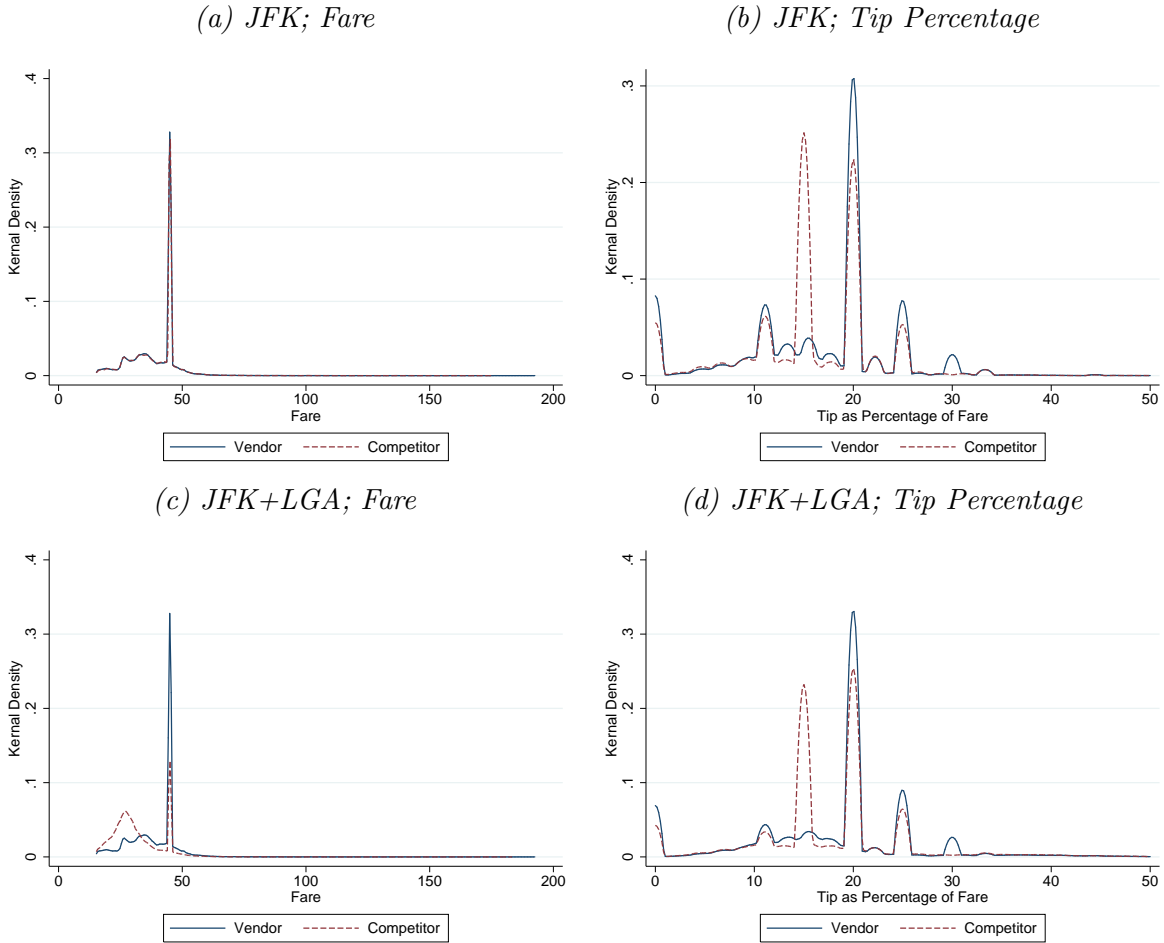
Figure C.2 is analogous to Figure 7 in the paper. The first row displays the Fare and Tip Percentage densities for the sample that limits rides that originate at JFK airport. The second row shows the pooled sample of rides originating at either LaGuardia or JFK airports. Finally, Table C.1 repeats Table 4 with this pooled sample of LGA and JFK rides, including a binary indicator for whether the ride originated at JFK.

Figure A.3: Proportion of Rides Originating with a Vendor versus a Competitor Equipped Cab, By census tract of Pick-Up Location



Notes: Graph on the left displays all of NYC, while the graph on the right displays only Manhattan. The sample is limited to fares between \$5 and \$25 on **Vendor-equipped** cab rides without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday).

Figure A.4: Distribution of Fares (A & C) and Tip Percentages (B & D) across Vendor and Competitor Equipped Taxis in JFK Sample (A & B) or JFK & LGA Sample (C & D)



Notes: The sample is limited to fares greater than \$15 on cab rides that originated at the census tract associated with JFK airport (Panels A & B) or those rides that originate at either the census tract associated with JFK airport or LaGuardia airport (Panels C & D), without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday). Panels B & D are limited to rides with tip percentages less than 50%.

Table A.3: OLS - Comparison of Vendor (20%/25%/30%) and Competitor (15%/20%/25%)

| | Fare | | Tip Percent | | Default Tip | | TipPercent0to10 | | Zero Tip | | Tip25 | |
|----------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|---------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Vendor | 0.419*** (0.060) | 0.203* (0.122) | 0.559*** (0.074) | 0.593*** (0.176) | -0.102*** (0.003) | -0.092*** (0.010) | -0.003** (0.001) | -0.009** (0.004) | 0.031*** (0.001) | 0.028*** (0.004) | 0.031*** (0.001) | 0.033*** (0.005) |
| Pick-Up at JFK | 15.368*** (0.036) | 15.560*** (0.059) | -1.071*** (0.088) | -2.152*** (0.090) | -0.070*** (0.004) | -0.065*** (0.005) | -0.022*** (0.001) | -0.010*** (0.002) | -0.006*** (0.002) | 0.012*** (0.002) | -0.018*** (0.002) | -0.015*** (0.003) |
| N | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 | 173,251 |
| r2 | 0.364 | 0.586 | 0.003 | 0.195 | 0.013 | 0.156 | 0.001 | 0.140 | 0.004 | 0.139 | 0.003 | 0.143 |
| MeanDepVar | 31.669 | 31.669 | 17.590 | 17.590 | 0.621 | 0.621 | 0.064 | 0.064 | 0.047 | 0.047 | 0.073 | 0.073 |
| Fixed Effects? | | X | | X | | X | | X | | X | | X |
| PVal_FEvsNoFE | | 0.000 | | 0.000 | | 0.180 | | 0.000 | | 0.000 | | 0.322 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the driver level, in parentheses. Even columns include fixed effects for driver, pick-up hour, and drop-off borough. Pick-Up at JFK is a binary variable that takes on value 1 if the ride originated at JFK airport. The dependent variable in columns 5 and 6 (Default Tip) takes on value 1 if the customer selected one of the default tip suggestions (buttons). The dependent variable in columns 7 and 8 (Tip Percent > 0 & < 10) takes on value 1 if the tip is greater than 0% and less than 10% of the fare. The dependent variable in columns 9 and 10 (Zero Tip) takes on value 1 if the customer left zero credit card tip. The dependent variable in columns 11 and 12 (Tip Percent = 25) takes on value 1 if the customer selected the 25% tip button. DepVarMean is the mean of the dependent variable in the control group (rides on Competitor-equipped cabs). PVal_FEvsNoFE is the p-value from a Chow Test for the equality of coefficients across even and odd columns. The sample is limited to fares greater than \$15 on cab rides that originated at the census tract associated with LaGuardia or John F. Kennedy Airports, without tolls, taxes, or surcharges (January 1, 2009 - October 31, 2009; 6am - 4pm on Monday - Friday and 6am - 8pm on Saturday and Sunday).

A.4 Data Cleaning

Our final dataset was constructed by first performing a number of consistency checks and then removing data that appeared to be generated by electronic tests or other types of data errors. The full sample of 170,896,479 was reduced to 13,820,735 observations by performing a number of procedures. We made the following consistency adjustments:

1. The pick-up came after the drop-off time in 0.14% (241,964) of observations. We replaced these pick-up times with their drop-off times, and vice-versa.
2. The drop-off time came after the pick-up time of the subsequent trip in 0.36% (618,570) of observations. We set the drop-off time equal to the pick-up time of the subsequent trip for all of these cases.

The full sample of 170,896,479 was reduced to 163,348,600 by dropping all observations for which:

1. There was a duplicate observation in terms of all original variables (750; 0.0004%).
2. The payment type was “No Charge” (509,194; .30%) or “Dispute” (94,784; .06%).
3. The ride duration was either equal to zero or longer than 3 hours (619,604; 0.36%).
4. The distance was either equal to zero or greater than 100 miles (929,498; 0.55%).
5. The surcharge was greater than \$1 (75,295; 0.04%).
6. Corresponding to drivers that drove fewer than 100 rides in 2009 (58,495; 0.03%).
7. Multiple cars were associated with the same driver during the same shift (1,298,412; 0.77%).

8. The driver's shift was longer than 20 hours (3,872,241; 2.31%).

9. The driver's shift was shorter than 30 minutes (89,606; 0.05%).

We then dropped the 5.95% (9,718,999) of observations that were on cars equipped with the third credit card machine vendor. Next, we dropped observations for which either the pick-up location or drop-off location could not be mapped to a census tract in NY, NJ, CT, or PA (2,022,218; 0.13%). To ensure that the regression discontinuity is identified off representative rides, we dropped all rides that had toll amounts applied. This dropped the 4,882,731 (3.22%) rides which were associated with a toll amount greater than zero. We then made the largest sample reduction, removing the 108,620,194 rides paid by cash, as the data did not include tip information for these rides. From this sample of 38,104,458 rides, we further limited to those rides for which the base amount (the sum of the fare, tolls, surcharge, and tax) was equivalent to the "fare". Performing this reduction ensured that rides on either side of the discontinuity were comparable in terms of the time of day, time of year, and the fees faced by the customer. This reduction left 13,929,933 rides that occurred prior to November 1, 2009 and between the hours of 6am to 4pm on Monday through Friday or 6am to 8pm on Saturday and Sunday. Finally, we removed rides that didn't correspond to a multiple of \$0.40 (the unit of fare accrual) added to \$2.50 (the flat entry fee), leaving 13,820,784 observations in the final sample.

Appendix B

Appendix for Chapter 2

B.1 (A)Data Cleaning

Our final dataset was constructed by first performing a number of consistency checks and then removing data that appeared to be generated by electronic tests or other types of data errors. Starting with 170,896,479, we performed the following data consistency adjustments:

1. The pick-up came after the drop-off time in 0.14% (241,965) of observations. We replaced these pick-up times with their drop-off times, and vice-versa.
2. The drop-off time came after the pick-up time of the subsequent trip in 0.36% (618,945) of observations. We set the drop-off time equal to the pick-up time of the subsequent trip for all of these cases.

We then classified a “Shift” as a succession of rides ending if the lag between two consecutive rides (the difference between drop-off time of the current ride and pick-up time of the subsequent ride) is longer than 300 minutes. This led to our next sample reductions:

1. Corresponding to drivers that drove fewer than 100 rides in 2009 (5,189 drivers; 12.58%).
2. Multiple cars were associated with the same driver during the same shift (1,298,409; 0.77%).
3. The driver's shift was longer than 20 hours (74,671 shifts; 0.99%).
4. The driver's shift was shorter than 30 minutes (36,986 shifts; 0.49%).

B.2 (B) Alternative Fraud Definitions

Our main fraud identification procedure corresponds to the “Maximum Fare Calculation” outlined in a May 14, 2010 TLC press release. They found 280,000 cases of fraud using 26 months of data and two methods to identify fraud:

- Time-Stamping: *A third of the available data includes a time-stamping of the Rate Code 4 activation. TLC screened out trips where the time-stamp indicates that Rate Code 4 was activated for less than 20% of total trip time.*
- Maximum Fare Calculation: *Two thirds of the available data have no time stamp available. Using data from the taxi technology systems in each taxicab, TLC captured the distance travelled and time elapsed during each trip where Rate Code 4 was activated while the taxi was in New York City. TLC calculated the maximum possible fare by adding 40 cents per minute (using an assumption that the cab was sitting in traffic for the entire trip) plus 40 cents per 1/5 mile travelled (assuming that the cab was moving above 12 miles-per-hour for the entire trip). At any given moment, a taxicab is either in traffic – and thus accruing “time” charges – or moving – and thus accruing “dis-*

tance” charges. So an actual fare will always be less than the maximum fare – as long as the proper rate is applied.

The 2009 data provided by the TLC does not provide a field for Rate Code, so we only classify rides as fraud if also start and end in Manhattan and could not have corresponded to the JFK rate.

There are a number of data issues that lead us to add a few additional conditions in order to identify a fare as fraudulent. In particular, there are cases in which either distance or duration are equal to zero, accounting for 672,330 of the 1,438,625 of rides classified as fraud in our first pass.

Table B.1: Fraud Summary Stats

| | (1) | (2) |
|---------------------------|------------|----------|
| | Fraud==0 | Fraud==1 |
| | mean/sd | mean/sd |
| Fare | (5.5085) | (9.8317) |
| Pr(Fraud) | 0.0000 | 1.0000 |
| | (0.0000) | (0.0000) |
| Fare Predicted | 9.2499 | 6.8113 |
| | (5.4384) | (4.5964) |
| Fare/Fare Predicted | 1.0075 | 1.7880 |
| | (0.0773) | (1.4910) |
| Fare minus Fare Predicted | 0.0407 | 5.2639 |
| | (0.8883) | (6.8604) |
| N | 69,014,761 | 611,019 |

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

Our correspondence with one of the largest two machine vendors suggests that these zero-valued entries are errors due to one of two possible reasons: (1) Service/training needs: Drivers and service technicians will run a trip to make sure the equipment works (generating a zero- distance and/or time trip). (2) For JFK and negotiated flat fares outside of Manhattan, drivers are supposed to engage the meter for the full trip, but some fail to engage the meter

until the end of the trip.

Furthermore, even after excluding these zero valued cases, 94,993 of the remaining rides classified as fraud have fares equal to \$45, the JFK rate. While these latter cases may have been legitimate cases of fraud, it is possible that drivers returning from JFK failed to turn on their meters until they were outside of the census tract corresponding to JFK (thus allowing the GPS to record a pick-up location outside of the JFK census tract). These latter instances have to be cases in which the meter was turned on after exiting the JFK census tract (or turned off prior to arriving at JFK), but for which the meter was activated during some portion of the trip (otherwise, these cases would have been removed in the step that excluded cases for which either distance or duration is equal to zero).

In consideration of these issues, our preferred fraud definition sets fraud equal to zero for rides in which either distance or duration are equal to zero, as well as all cases in which the fare is equal to \$45.

We present LPM regressions based on three further fraud definitions. These further fraud variables were constructed as follows:

1. Fraud Definition 2: Allowing for JFK Rate Fraud: Again, we set fraud equal to zero for cases in which either duration or distance is missing; however, we allow for fraud to be equal to one for the remaining cases in which fare = \$45. While some of these may correspond to driver errors, it is possible that these were legitimate cases of fraud. This definition results in 766,295 cases of fraud.
2. Fraud Definition 3: Imputation for Missing Values: As with Definition 1 we exclude cases in which both distance *and* duration are missing; however, we impute conservative values for cases in which only one of the two measures is equal to zero. We rely on the assumption that if only one of the two measures equals zero, the other measure

is still accurate. If distance is equal to zero, we set it equal to duration * 30 (i.e. we assume an average driving speed of 2mph). If duration is equal to zero, we set it equal to distance (i.e. we assume an average driving speed of 60mph). As with the main definition, we set fraud equal to zero if the fare is equal to \$45. This definition results in 717,545 cases of fraud.

3. Fraud Definition 4: Trip Time: We follow the same rules used in Definition 1; however, we use a different variable for duration. The data set provides a rounded “trip_time” variable in addition to the pick-up and drop-off times. In our previous three measures, we defined duration as the difference between pick-up and drop-off time, allowing for a measurement in seconds.¹ In most cases, the “trip_time” is equal to this difference, rounded to the nearest integer (minutes). However, there are several cases in which the constructed duration variable equals zero and “trip_time” is non-zero, and vice versa. Using “trip_time” as our duration variable yields 929,361 cases of fraud.

The following table presents fraud summary statistics for these alternative fraud definitions.

¹We round this measure up to the next highest minute integer, following a conservative reading of the rule listed on the TLC website, “The fare shall include pre-assessment of the unit currently being accrued; the amount due may therefore include a full unit charge for a final, fractional unit.” In cleaning the data, we also flipped pick-up time and drop-off time if pick-up time came after drop-off time.

Table B.2: Other Fraud Definitions - Summary Stats

| | (1) | (2) | (3) | (4) |
|-----------------|-----------|-----------|-----------|-----------|
| | Trip=1 | Trip=5 | Trip=20 | Last Trip |
| | mean/sd | mean/sd | mean/sd | mean/sd |
| Pr(Fraud) | 0.0109 | 0.0101 | 0.0076 | 0.0103 |
| | (0.1039) | (0.0998) | (0.0869) | (0.1011) |
| Pr(Fraud def.2) | 0.0109 | 0.0101 | 0.0076 | 0.0103 |
| | (0.1039) | (0.0998) | (0.0869) | (0.1011) |
| Pr(Fraud def.3) | 0.0136 | 0.0131 | 0.0094 | 0.0123 |
| | (0.1160) | (0.1135) | (0.0966) | (0.1103) |
| Pr(Fraud def.4) | 0.0109 | 0.0101 | 0.0076 | 0.0103 |
| | (0.1039) | (0.0998) | (0.0869) | (0.1011) |
| N | 2,726,856 | 2,867,509 | 2,001,056 | 2,820,122 |

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

We then present results from linear probability model regressions that control for shift fixed effects for our four fraud definitions, replicating the specification presented in Table 3.

Table B.3: Other Fraud Definitions - LPM

| | (1) | (2) | (3) | (4) |
|--|-----------------------|-----------------------|-----------------------|-----------------------|
| | Fraud | Fraud def.2 | Fraud def. 3 | Fraud def. 4 |
| | b/se | b/se | b/se | b/se |
| $\mathbf{1}(\text{Gap}(i,s,r) \geq 0)$ | 0.0006*** (0.0000) | 0.0006*** (0.0000) | 0.0016*** (0.0000) | 0.0006*** (0.0000) |
| Pr(Fraud) | 0.0088 | 0.0088 | 0.0088 | 0.0088 |
| Observations | 69,625,780 | 69,625,780 | 69,625,780 | 69,625,780 |
| r2 | 0.4037 | 0.4037 | 0.4066 | 0.4037 |
| Shifts | 3,036,570 | 3,036,570 | 3,036,570 | 3,036,570 |
| Drivers | 12,451 | 12,451 | 12,451 | 12,451 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. The sample is limited to a 30% random sample of drivers who committed fraud (definition 1) at least 10 times in 2009, we selected all drivers for whom we record an instance of fraud at least 10 times in 2009 (3798 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature $> 80F$, temperature $< 30F$), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

B.3 (C)Robustness

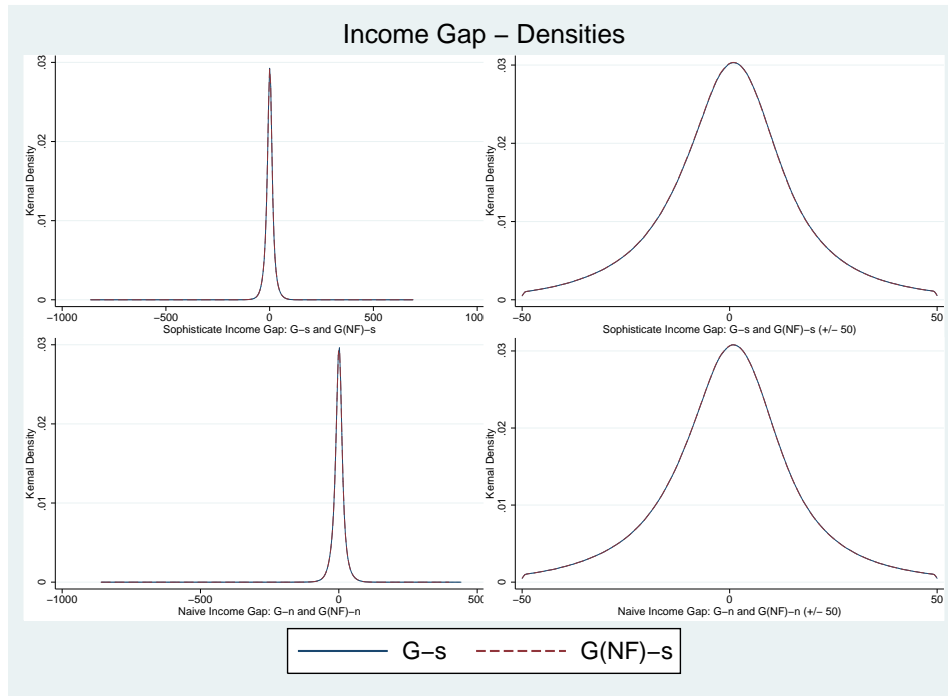
B.3.1 Alternative Income Gaps

In the paper, we present results for a measure of the Income Gap that does not include frauds and includes a measure of the predicted fare on the current ride in the gap. In this section, we show that results are robust to using a gap constructed on income on previous rides only and to the inclusion of frauds. We construct a “naive” $\text{Gap}(i, s, r)[II] = \hat{I}_{i,s,r-1} - I_{i,s,r-1}$ as

a measure that does not include the perspective fare. In this measure, a driver compares the amount of money usually earned, $\hat{I}_{i,s,r-1}$, with the amount of money earned on the current shift up to the previous ride, $I_{i,s,r-1}$. We then present two further measures. The first $Gap(i, s, r)[III]$ is constructed as our main Income Gap measure, but it includes income earned on cheated rides. The second, $Gap(i, s, r)[IV]$ mimic our naive gap measure, but it also includes income from frauds.

First, we show that the estimated gap are normally distributed around zero, and that the inclusion of the predicted fare does not generate large differences among the naive and sophisticated gap measures.

Figure B.1: Alternative Income Gap Measures



Notes: Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

We then present summary statistics on the other gaps.

Table B.4: Summary Statistics: Alternative Income Gap Measures

| | (1) | (2) | (3) | (4) | (5) |
|------------------------|---------------------|----------------------|-----------------------|-----------------------|-----------------------|
| | Trip=1 | Trip=5 | Trip=20 | Last Trip | All |
| | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd |
| Cumulative (NF) Income | 10.9961 (7.2634) | 51.3552 (22.0021) | 184.5148 (35.6113) | 233.5395 (84.3129) | 133.8621 (88.7124) |
| Expected (NF) Income | 14.2662 (9.0737) | 52.4539 (22.1111) | 185.0658 (30.5446) | 233.7426 (76.1249) | 134.8260 (85.4284) |
| Gap(i,s,r)[II] | . | 1.0016 (13.8588) | 0.6431 (22.0828) | 0.2539 (28.5334) | 0.1584 (19.1835) |
| Cumulative Income | 11.0744 (7.3615) | 51.6099 (22.1434) | 185.2094 (35.6095) | 234.5656 (84.4072) | 134.3920 (89.0114) |
| Expected Income | 14.3233 (9.1449) | 52.7216 (22.2581) | 185.7739 (30.6010) | 234.7450 (76.2226) | 135.3511 (85.7277) |
| Gap(i,s,r)[III] | 2.9901 (8.3396) | 1.3016 (14.6519) | 0.7760 (22.6767) | 0.1635 (30.0940) | 0.4222 (20.1322) |
| Gap(i,s,r)[IV] | . | 1.0123 (13.9269) | 0.6390 (22.1074) | 0.2381 (28.5892) | 0.1573 (19.2350) |
| N | 2,726,856 | 2,867,509 | 2,001,056 | 2,820,122 | 69,625,780 |

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

Below, we present regressions from a linear probability model that compares among the

four measures. All regressions include driver-shift fixed effects.

Table B.5: LPM for Alternative Income Gap Measures

| | (1) | (2) | (3) | (4) |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| | Fraud | Fraud | Fraud | Fraud |
| | b/se | b/se | b/se | b/se |
| $1(\text{Gap}(i,s,r) \geq 0)$ | 0.0006*** (0.0000) | | | |
| $1(\text{Gap}(i,s,r)[\text{II}] \geq 0)$ | | 0.0002*** (0.0000) | | |
| $1(\text{Gap}(i,s,r)[\text{III}] \geq 0)$ | | | 0.0021*** (0.0000) | |
| $1(\text{Gap}(i,s,r)[\text{IV}] \geq 0)$ | | | | 0.0013*** (0.0000) |
| Pr(Fraud) | 0.0088 | 0.0088 | 0.0088 | 0.0088 |
| Observations | 69,625,780 | 69,625,780 | 69,625,780 | 69,625,780 |
| r ² | 0.4037 | 0.4037 | 0.4037 | 0.4037 |
| Shifts | 3,036,570 | 3,036,570 | 3,036,570 | 3,036,570 |
| Drivers | 12,451 | 12,451 | 12,451 | 12,451 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

Finally, we present regressions from a linear probability model that compares among measures of the income gap that included different orders of the polynomial in minutes elapsed since the beginning of the shift. All regressions include driver-shift fixed effects. The

coefficients are unchanged up to the fourth digit.

Table B.6: LPM for Polynomial Income Gap Measures

| | (1) | (2) | (3) | (4) |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| | Fraud | Fraud | Fraud | Fraud |
| | b/se | b/se | b/se | b/se |
| $\mathbb{1}(\text{Gap}(i,s,r)[1st] \geq 0)$ | 0.0006*** (0.0000) | | | |
| $\mathbb{1}(\text{Gap}(i,s,r) \geq 0)$ | | 0.0006*** (0.0000) | | |
| $\mathbb{1}(\text{Gap}(i,s,r)[3rd] \geq 0)$ | | | 0.0006*** (0.0000) | |
| $\mathbb{1}(\text{Gap}(i,s,r)[4th] \geq 0)$ | | | | 0.0006*** (0.0000) |
| Pr(Fraud) | 0.0084 | 0.0084 | 0.0084 | 0.0084 |
| Observations | 73,033,879 | 73,033,879 | 73,033,879 | 73,033,879 |
| r ² | 0.3778 | 0.3778 | 0.3778 | 0.3778 |
| Shifts | 3,074,836 | 3,074,836 | 3,074,836 | 3,074,836 |
| Drivers | 12,446 | 12,446 | 12,446 | 12,446 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

B.3.2 Autocorrelation

Here we show the autocorrelation structure of our measure of the Income Gap. We find, if anything, a slightly negative correlation.

Table B.7: LPM with income inclusive frauds - Panel B

| | (1) | (2) | (3) |
|--------------------------------|------------------------|------------------------|--------------------------------|
| | Shift Income | Income up to Hour 6 | $mean(Gap(i, s, r))$ in Hour 6 |
| | b/se | b/se | b/se |
| $mean(Gap(i, s, r))$ in Hour 6 | -1.2656*** (0.0064) | -1.3158*** (0.0025) | |
| $mean(Gap(i, s, r))$ in Hour 4 | 0.0134* (0.0074) | 0.0570*** (0.0026) | 1.1232*** (0.0015) |
| $mean(Gap(i, s, r))$ in Hour 2 | 0.5561*** (0.0107) | 0.4918*** (0.0048) | -0.1985*** (0.0021) |
| Observations | 1,717,010 | 1,717,010 | 1,717,010 |
| r ² | 0.5333 | 0.7655 | 0.6786 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature $> 80F$, temperature $< 30F$), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

B.3.3 Fixed effects contributions to Shift Income and Frauds

Here we present the fixed effects coefficients for our main regressions.

Table B.8: LPM with income inclusive frauds - Panel B

| | (1) | (2) |
|--------------|-------------------------|-------------------------|
| | shift_income | shift_income |
| | b/se | b/se |
| Midtown | -2.4067*** (0.1422) | -5.9356*** (0.2976) |
| Upper-West | 5.5004*** (0.1561) | 1.9762*** (0.3891) |
| Upper-East | 0.4314** (0.2105) | 0.2247 (0.5096) |
| Uptown | 10.0534*** (0.2027) | 7.0808*** (0.5066) |
| Brooklyn | 11.2184*** (0.7326) | 4.5434*** (0.9411) |
| Bronx | -3.9385*** (0.2145) | 0.8397** (0.4262) |
| Queens | 0.4680** (0.2261) | 4.8299*** (0.4313) |
| JFK | -9.5693*** (0.2214) | -0.5485 (0.6060) |
| La Guardia | -3.9189*** (0.1904) | -0.1269 (0.5677) |
| Others | -2.9711*** (0.3842) | 0.4323 (0.6234) |
| Daily Shift | 20.3187*** (0.0950) | 43.5779*** (0.5898) |
| Monday | -1.3755*** (0.1908) | -1.8256*** (0.5719) |
| Tuesday | 12.0041*** (0.1876) | 13.0616*** (0.5776) |
| Wednesday | 18.5237*** (0.1872) | 20.0864*** (0.5805) |
| Thursday | 38.4131*** (0.1858) | 39.8438*** (0.6080) |
| Friday | 67.6285*** (0.1871) | 69.6371*** (0.6992) |
| Saturday | 57.5401*** (0.1909) | 59.0188*** (0.7501) |
| February | 12.4148*** (0.2203) | 12.3045*** (0.2337) |
| March | 10.1090*** (0.2158) | 9.9414*** (0.2716) |
| April | 9.5593*** (0.2184) | 9.5894*** (0.2881) |
| May | 9.2902*** (0.2184) | 9.1897*** (0.3005) |
| June | 7.4295*** (0.2205) | 7.4468*** (0.2142) |
| July | -5.8012*** (0.2210) | -5.7632*** (0.3246) |
| August | -4.3703*** (0.2226) | -4.3823*** (0.3360) |
| September | 9.1920*** (0.2229) | 8.7162*** (0.3494) |
| October | 17.3316*** (0.2192) | 16.7417*** (0.3560) |
| November | 8.3717*** (0.2233) | 7.3504*** (0.3629) |
| December | 5.1791*** (0.2208) | 4.2763*** (0.3680) |
| Constant | 189.2073*** (0.2308) | 179.3232*** (0.7238) |
| Observations | 3,075,298 | 3,075,298 |
| r2 | 0.1065 | 0.3730 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

Table B.9: LPM with income inclusive frauds - Panel B

| | (1) | (2) | (3) |
|----------------------------|------------------------|------------------------|------------------------|
| | Fraud | Fraud | Fraud |
| | b/se | b/se | b/se |
| Distance | (0.0000) | (0.0000) | (0.0000) |
| Duration | (0.0000) | (0.0000) | (0.0000) |
| Temperature is below 30F | 0.0123*** (0.0002) | 0.0130*** (0.0002) | -0.0007*** (0.0001) |
| Temperature is above 80F | -0.0017*** (0.0001) | -0.0021*** (0.0001) | -0.0022*** (0.0001) |
| Rain ever today | -0.0004*** (0.0001) | -0.0002*** (0.0000) | -0.0001*** (0.0001) |
| (max LGA/JFK/Central Park) | (0.0001) | (0.0000) | (0.0001) |
| 1h.Midtown | (.) | (.) | (.) |
| 2.Uptown | (0.0000) | (0.0000) | (0.0000) |
| 3.Bronx | (0.0000) | (0.0000) | (0.0000) |
| 4.Queens | (0.0003) | (0.0002) | (0.0002) |
| 5.Brooklyn | (0.0001) | (0.0001) | (0.0001) |
| 6.JFK | (0.0001) | (0.0001) | (0.0000) |
| 8.La Guardia | (0.0003) | (0.0002) | (0.0002) |
| 9.Others | (0.0016) | (0.0011) | (0.0011) |
| Hour 2 | 0.0010*** (0.0001) | 0.0007*** (0.0001) | 0.0008*** (0.0001) |
| Hour 3 | 0.0008*** (0.0001) | 0.0004*** (0.0001) | 0.0006*** (0.0001) |
| Hour 4 | 0.0008*** (0.0001) | 0.0007*** (0.0001) | 0.0010*** (0.0001) |
| Hour 5 | 0.0050*** (0.0001) | 0.0044*** (0.0001) | 0.0032*** (0.0001) |
| Hour 6 | 0.0071*** (0.0002) | 0.0019*** (0.0002) | 0.0012*** (0.0002) |
| Hour 7 | 0.0057*** (0.0002) | -0.0024*** (0.0002) | -0.0021*** (0.0001) |
| Hour 8 | 0.0033*** (0.0002) | -0.0036*** (0.0001) | -0.0027*** (0.0001) |
| Hour 9 | 0.0010*** (0.0001) | -0.0043*** (0.0001) | -0.0027*** (0.0001) |
| Hour 10 | 0.0006*** (0.0001) | -0.0043*** (0.0001) | -0.0022*** (0.0001) |
| Hour 11 | 0.0007*** (0.0001) | -0.0044*** (0.0001) | -0.0024*** (0.0001) |
| Hour 12 | 0.0007*** (0.0001) | -0.0044*** (0.0001) | -0.0022*** (0.0001) |
| Hour 13 | 0.0002 (0.0001) | -0.0047*** (0.0001) | -0.0025*** (0.0001) |
| Hour 14 | 0.0005*** (0.0001) | -0.0041*** (0.0001) | -0.0021*** (0.0001) |
| Hour 15 | -0.0004*** (0.0001) | -0.0028*** (0.0001) | -0.0019*** (0.0001) |
| Hour 16 | -0.0012*** (0.0001) | -0.0030*** (0.0001) | -0.0013*** (0.0001) |
| Hour 17 | 0.0032*** (0.0001) | 0.0030*** (0.0001) | 0.0041*** (0.0001) |
| Hour 18 | 0.0022*** (0.0001) | 0.0034*** (0.0001) | 0.0043*** (0.0001) |
| Hour 19 | 0.0009*** (0.0001) | 0.0031*** (0.0001) | 0.0039*** (0.0001) |
| Hour 20 | 0.0013*** (0.0001) | 0.0030*** (0.0001) | 0.0034*** (0.0001) |
| Hour 21 | -0.0008*** (0.0001) | 0.0001*** (0.0000) | 0.0002*** (0.0000) |
| Hour 22 | -0.0003*** (0.0001) | 0.0003*** (0.0000) | 0.0003*** (0.0000) |
| Hour 23 | 0.0001*** (0.0001) | 0.0006*** (0.0000) | 0.0006*** (0.0000) |
| Hour 24 | 0.0005*** (0.0001) | 0.0008*** (0.0000) | 0.0009*** (0.0000) |
| Constant | 0.0088*** (0.0001) | 0.0095*** (0.0000) | 0.0098*** (0.0000) |
| Observations | 69,625,780 | 69,625,780 | 69,625,780 |
| r ² | 0.0001 | 0.2428 | 0.4036 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

B.3.4 JFK/LGA regression table

As explained in the paper, when a driver drops a customer off at JFK or LGA, he has a choice of whether to go back empty to the city or wait at the airport. Drivers who decide to stay must queue for a variable time period that depends on the intersection of demand (travelers) and supply (other drivers ahead in the queue). The Port Authority allocates the drivers to the several terminals to meet new customers. The queues of both passengers and drivers produce quasi-random variation in the match characteristics. We exploit this double source of variation in waiting time and passenger characteristics, to estimate the results reported in the Table below. Specifically, we use the continuous variable of waiting time in minutes to instrument for an indicator that the driver is above target. We restrict the sample to the rides after the one starting at JFK or LGA (on this ride, frauds are excluded in our definition). Because, in most shifts, drivers are observed at JFK/LGA only once, these regressions include driver fixed effects, so as not to limit variation to the handful of rides kept within each shift. The table presents instrumental variable regressions using the waiting time at JFK or LGA (columns 1 and 3) and the waiting time of other drivers over the same hour (columns 2 and 4) as instruments. Moreover, the first two columns report results on the first ride after the ride from JFK (this latter is excluded from our fraud definition), while columns 3 and 4 report results on all subsequent rides. The regression further controls for seasonality, which was accounted for in our measure of the Income Gap, by including month of the year fixed effects. Only in regressions including all rides after the airport's stay, we do find significant coefficients consistent with our main analysis. Restricting to the second ride after JFK or LGA we do not find a result.

Table B.10: IV: 2SLS Estimates and First-Stage (JFK)

| | Second Ride After JFK/LGA | | All Rides After JFK/LGA | |
|---|---------------------------|-----------|-------------------------|-----------|
| | (1) | (2) | (3) | (4) |
| IV Estimates - DepVar: Fraud | | | | |
| 1(Gap(i,s,r) ≥ 0) | -0.0015 | -0.0019 | 0.0025** | 0.0037** |
| | (0.0017) | (0.0028) | (0.0012) | (0.0018) |
| First Stage - DepVar: 1(Gap(i,s,r) ≥ 0) | | | | |
| Wait Time (airport) | 0.0029*** | | 0.0024*** | |
| | (0.0000) | | (0.0000) | |
| Others Wait Time (airport) | | 0.0035*** | | 0.0031*** |
| | | (0.0000) | | (0.0000) |
| N | 364,048 | 376,109 | 4,969,936 | 5,131,029 |
| r2 | 0.3904 | 0.3503 | 0.2133 | 0.1923 |
| Shifts | 363,817 | 363,817 | 363,817 | 363,817 |
| Drivers | 12,103 | 12,103 | 12,103 | 12,103 |
| Fixed Effects | Driver | Driver | Driver | Driver |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at the shift-driver level, in parentheses. Data restricted to drivers for whom we record an instance of fraud at least 10 times in 2009 (12451 drivers). Rides limited to those in which it is possible to detect fraud: those starting and ending in New York City (excluding JFK); non-zero duration and distance, non-\$45 fare. All specifications include dummies for weather (rain, temperature > 80F, temperature < 30F), location (downtown, uptown, Bronx, Queens, Brooklyn, LaGuardia Airport), and hour of the day.

Appendix C

Appendix for Chapter 3

C.1 Summary Statistics

In this Appendix, summary statistics for full sample by year and round of the program are reported.

Table C.1: Summary statistics by program round and year

| | Investment Program Round 2 | | | | | | | | |
|--------------------------|----------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|---------------------|--------------------|--------------------|
| | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | Total |
| | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd |
| Investments/Sales | . (.) | .1226 (.2318) | .1051 (.1775) | .1399 (.2268) | .13 (.2162) | .1129 (.2119) | .1287 (.2454) | .08151 (.1726) | .1173 (.2139) |
| Profit/Sales | .3228 (.1964) | .3206 (.1832) | .3268 (.1735) | .3176 (.1752) | .3099 (.1671) | .3114 (.1679) | .305 (.1777) | .3016 (.1688) | .3145 (.1766) |
| Labor Costs/Sales | . (.) | .1905 (.1402) | .1985 (.1435) | .2031 (.1368) | .1973 (.1249) | .2015 (.1298) | .1998 (.1387) | .2048 (.143) | .1994 (.1369) |
| Assets (in thousands) | 13,465 (60,961) | 14,847 (64,439) | 15,435 (66,037) | 16,663 (65,739) | 18,009 (71,353) | 20,354 (83,150) | 24,904 (143,400) | 22,587 (77,288) | 18,283 (83,030) |
| Debt (in thousands) | 8,168 (32,998) | 9,046 (35,698) | 9,033 (34,949) | 9,612 (33,675) | 10,453 (38,286) | 12,292 (48,690) | 15,785 (104,956) | 13,510 (43,371) | 10,987 (51,816) |
| Cash Flow (in thousands) | . (.) | 1,569 (7,442) | 1,563 (8,406) | 1,873 (9,796) | 1,816 (9,184) | 2,111 (11,702) | 3,872 (56,739) | 1,932 (10,227) | 2,105 (23,199) |
| Sales (in thousands) | 13,226 (58,486) | 15,948 (67,225) | 16,638 (72,124) | 17,832 (73,807) | 18,655 (73,605) | 19,535 (75,429) | 22,140 (79,874) | 22,867 (82,769) | 18,355 (73,277) |
| Investments/Capital | . (.) | 1.277 (8.753) | .8653 (2.882) | .9722 (2.94) | .6059 (2.764) | .3518 (.7212) | .4645 (1.143) | .2777 (.727) | .6878 (3.859) |
| Observations | 891 | 891 | 891 | 891 | 891 | 891 | 891 | 891 | |

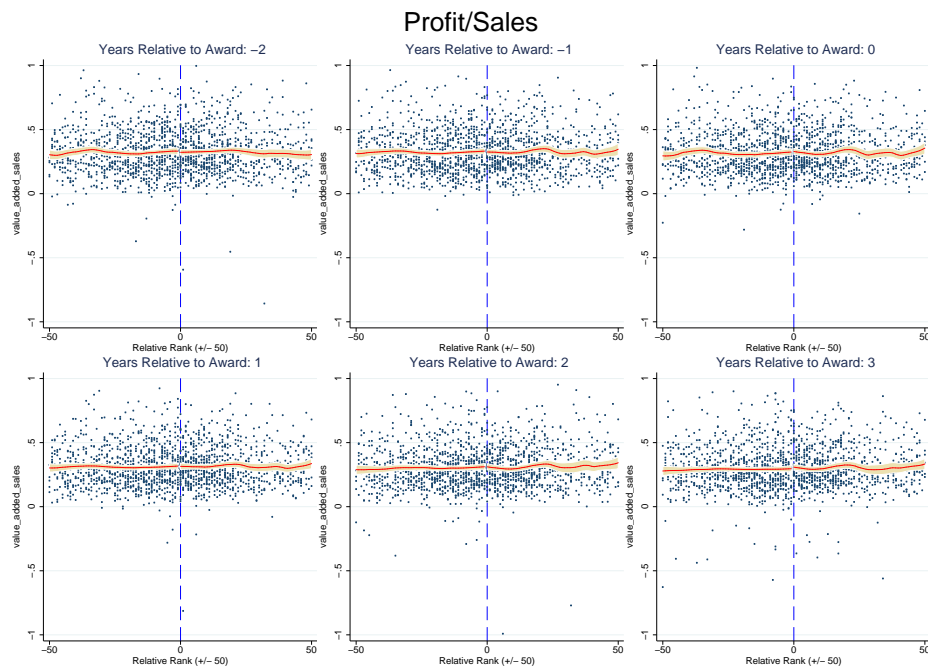
| | Investment Program Round 3 | | | | | | | |
|--------------------------|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | Total |
| | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd | mean/sd |
| Investments/Sales | . (.) | .09797 (.1845) | .1104 (.1872) | .1127 (.1923) | .09993 (.1643) | .1034 (.1856) | .07809 (.1696) | .1004 (.1812) |
| Profit/Sales | .3109 (.1827) | .3189 (.1771) | .3162 (.1665) | .3085 (.1653) | .3106 (.1639) | .2994 (.1707) | .2894 (.1806) | .3077 (.1727) |
| Labor Costs/Sales | .1811 (.1394) | .188 (.137) | .1918 (.1324) | .1885 (.1271) | .1942 (.1314) | .1891 (.1284) | .1942 (.1377) | .1896 (.1334) |
| Assets (in thousands) | 18,267 (126,267) | 18,180 (119,519) | 19,823 (127,714) | 21,252 (134,453) | 22,285 (137,767) | 25,016 (156,683) | 26,079 (161,730) | 21,557 (138,495) |
| Debt (in thousands) | 11,212 (77,728) | 11,050 (73,765) | 11,911 (77,848) | 12,356 (78,834) | 13,198 (81,930) | 14,511 (87,088) | 14,954 (90,479) | 12,742 (81,261) |
| Cash Flow (in thousands) | 1,885 (14,391) | 1,665 (13,709) | 2,037 (14,500) | 2,272 (14,940) | 2,618 (18,152) | 2,910 (23,000) | 2,675 (22,811) | 2,295 (17,757) |
| Sales (in thousands) | 17,867 (127,167) | 19,125 (132,508) | 20,761 (141,848) | 22,248 (156,007) | 22,844 (157,992) | 26,330 (179,289) | 26,952 (179,885) | 22,304 (154,738) |
| Investments/Capital | . (.) | .6869 (2.072) | 1.19 (9.739) | .8505 (5.772) | .6546 (2.584) | .6238 (3.848) | .3384 (.9535) | .7241 (5.085) |
| Observations | 1220 | 1220 | 1220 | 1220 | 1220 | 1220 | 1220 | 1220 |

Notes: The sample is limited to 891 firms for the second round and 1,220 firms for the third round.

C.2 Graphical evidence for other Outcome Variables

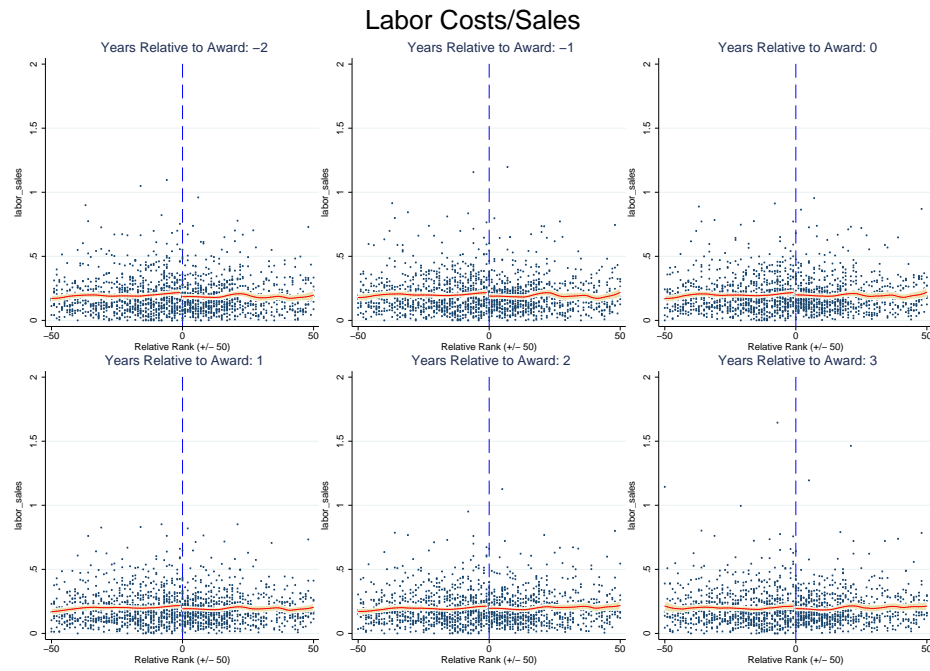
The following two figures report visual evidence for two outcome variables included in the regression analysis. The graphs are constructed using the methodology explained in the main text.

Figure C.1: Local polynomial regressions for profit/sales for three years before and after the award



Notes: Local polynomial regressions by rank position relative to cutoff. Standard errors in shaded areas. The dots represent means of the y-axis variable for each position in the relative rank.

Figure C.2: Local polynomial regressions for labor costs/sales for three years before and after the award



Notes: Local polynomial regressions by rank position relative to cutoff. Standard errors in shaded areas. The dots represent means of the y-axis variable for each position in the relative rank.